University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Black, David John

Title:
The utilisation of human capital from education in Australian labour markets: over-education?

Date:
2013

Citation:
Black, D. J. (2013). The utilisation of human capital from education in Australian labour markets: over-education?. PhD thesis, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Publication Status:
Unpublished

Persistent Link:
https://hdl.handle.net/11343/38342

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

The Utilisation of Human Capital from Education in Australian Labour Markets: Over-education?

David John Black

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy
(with coursework component)

March 2013

Melbourne Institute of Applied Economic and Social Research,
The University of Melbourne

(Produced on archival quality paper)

Abstract

Modern societies devote considerable resources to the education of individuals. This represents an investment for individuals and societies. Since many of the benefits are realised via the employment of individuals, labour markets are critical for the accrual of benefits to such investments in education. Labour market failures, therefore, can adversely affect the benefits that accrue to individuals and societies. In particular, their failure to facilitate the full utilisation of the human capital individuals derive from education—a phenomenon referred to as *over-education*—would diminish such benefits. For societies, over-education implies an under-utilisation of human capital available in the workforce, leading to productivity levels, economic growth rates and living standards below their potential. Meanwhile, over-educated individuals would not receive the full benefits to their investments, at least in terms of increased earnings. This study aims to investigate the existence of over-education in Australian labour markets. It uses individual-level data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey for the 2001 to 2008 period, where initial results indicate roughly 20 per cent of working-age employees are over-educated in each year.

The study considers three key questions regarding over-education. The first concerns its identification and subsequent assertion that individuals deemed over-educated have human capital from education that is under-utilised in their current job. This pivotal assertion is empirically tested. The results indicate those identified as over-educated do indeed have under-utilised human capital. The second question concerns the dynamics of over-education. In particular, the possibility that over-education is a result of the inherently dynamic nature of labour markets, whereby it may be merely short-term disequilibria with no enduring effects. While it is predominantly a temporary state, a significant proportion of affected individuals are found to be persistently over-educated. It is also found to have detrimental effects that endure beyond individuals' exits from the state; specifically, evidence indicates over-education can lead to human capital depreciation. Over-education, therefore, is not merely a by-product of adjustment processes in dynamic and well-functioning labour markets. The third question concerns the possibility not all over-education represents labour market failures. In particular, since many job attributes, not just the wage, can affect individuals' utility levels, some over-educated individuals may have obtained jobs that maximise their (expected) utility levels and, therefore, achieved their preferred outcome. Thus, individuals may trade wages for non-pecuniary benefits or improved working conditions and, thereby, accept jobs for which they are over-educated. The results indicate roughly one-third of over-educated individuals are actually *voluntarily over-educated*. They are found to trade wages for increased job security, preferred hours, greater job flexibility and reduced stress. And, overall, they are more satisfied with their achieved work-life balance.

Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Also, this study uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this study, however, are those of the author and should not be attributed to either FaHCSIA or the Melbourne Institute.

David J. Black

Acknowledgements

The following is dedicated to my beautiful wife, Angela. Without her this would have been impossible for so many reasons. Her years of sacrifice and support gave me the opportunity to pursue such an endeavour, but, most importantly, it was her love and unwavering belief in me that provided the inspiration, motivation, resilience and self-belief necessary to complete this PhD thesis. Her continued efforts at patience and understanding during the final months were also noticed and greatly appreciated—though I did benefit from the occasional *kick-in-the-pants*. AC, I love you and I sincerely thank you for helping me make this dream a reality. Thank you also to my parents, Keith and Marie. Their love and support, along with many years of sacrifices and encouragement, helped me reach the point where this endeavour was possible. They provided me with a truly privileged life. Any and all of my successes are thanks to these three people—I will be forever indebted to each of them.

While all subsequent research is my own, this thesis is also the result of the hard work and dedication of my two supervisors, Jeff Borland and Mark Wooden. Their guidance, feedback and patience have been invaluable and greatly appreciated. And this thesis is immeasurably better for their contributions. I am extremely grateful to have had the opportunity to work with two of the finest economists and academics in Australia. Thank you both.

I have also had the benefit of learning from colleagues at the Melbourne Institute. I thank Yi-Ping Tseng and Roger Wilkins for teaching me the art of data analysis using Stata, for encouraging me to complete a PhD and for always treating me as a colleague and co-author rather than research assistant. I thank Bob Drago and Mark Wooden for showing me the process of publishing in academic journals, and for my first taste of publication success. Thanks also to Paul Jensen, Umut Oguzoglu, Wang-Sheng Lee, Guyonne Kalb, Kostas Mavromaras and Cain Polidano for teaching me various aspects of the research process. It has been a privilege to work with such intelligent, dedicated and kind individuals. Thank you all.

Thanks to the Melbourne Institute for providing financial support and an environment conducive to research. And thanks to past students Rosanna Scutella, Ha Vu and Diana Warren for their example, advice and encouragement in completing a PhD thesis. Thanks also to David Evans and Paul Gruba for the excellent guidance provided in their book *How to Write a Better Thesis*.

Finally, thanks to my friends and family for their moral support, especially those who recognised my reluctance to discuss this thesis and promptly changed the subject. Special thanks to Mark Burkhart and Daniel Kuehnle for always being willing to listen to me whine over a few beers. And thanks to Vader Black, my cat, for companionship on late nights and weekends spent working.

Contents

Chapter 1

Introduction 1

Chapter 2

Over-education research 8

2.1 Introduction 8
2.2 Theoretical foundation for over-education 9
2.3 Empirical identification of over-education 15
2.3.1 Methods for estimating required education levels of jobs 17
2.4 Over-education wage penalty estimates 24
2.5 Dynamics of over-education 28
2.5.1 Persistence for individuals 29
2.5.2 Link to human capital depreciation 31
2.6 Individual preferences and over-education 32
2.7 Over-education in Australia 36
2.8 Conclusion 37

Chapter 3

Research design 39

3.1 Introduction 39
3.2 Research questions 39
3.3 Research method – Empirical tests of the research questions 41
3.3.1 Over-education and under-utilised human capital 44
3.3.2 Dynamics of over-education 45
3.3.3 Voluntary over-education 47
3.4 The HILDA Survey data 49
3.4.1 Education data 50
3.4.2 Occupation data 54
3.4.3 Sample attrition 61
3.5 Method used to identify over-education 65

Chapter 4

Incidence of over-education in Australia 71

4.1 Introduction 71
4.2 Estimated incidence of over-education 71
4.3 Comparison with previous estimates 74
4.4 Discussion and conclusion 76

Chapter 5

Over-education and under-utilised human capital 77

5.1 Introduction 77
5.2 Existing evidence 79
5.3 Framework and empirical estimators 81
5.3.1 Causal effect of over-education 81

5.3.2 Parametric estimators – Regression estimates of ORU earnings functions 82
5.3.3 Semi-parametric estimators – Matching estimates 85
5.4 Empirical results 89
5.4.1 Descriptive statistics 90
5.4.2 Regression estimates 93
5.4.3 Cross-sectional matching estimates 97
5.4.4 Combined matching and regression estimates 101
5.4.5 Difference-in-differences matching estimates 102
5.5 Synthesis of results 105
5.6 Discussion and conclusion 107

Chapter 6

Dynamics of over-education 109
6.1 Introduction 109
6.2 Existing evidence 111
6.3 Persistence of over-education 112
6.3.1 Transition rates and durations over-educated 112
6.3.2 Wage penalty by duration over-educated 116
6.4 State dependence in over-education 122
6.4.1 Econometric model and estimators 124
6.4.2 Empirical results 128
6.5 Prior over-education and wages of well-matched individuals 134
6.5.1 Econometric model and estimators 135
6.5.2 Empirical results 135
6.6 Discussion and conclusion 140

Chapter 7

Voluntary over-education 143
7.1 Introduction 143
7.2 Existing evidence 144
7.3 Estimating the incidence of voluntary over-education 145
7.4 Voluntary over-education and job attributes 149
7.4.1 Econometric models and estimators 150
7.4.2 Empirical results 155
7.5 Persistence and voluntary over-education 164
7.6 Discussion and conclusion 165

Chapter 8

Conclusions 169

References 173

Appendices

Appendix 1.1: Key definitions, assumptions and limitations of study 188
Appendix 2.1: Alternative theoretical frameworks for over-education 190
Appendix 3.1: Details on the structure of ANZSCO 194
Appendix 3.2: Alterations to highest education level variables 196
Appendix 3.3: Benchmarking measures of highest education level 197
Appendix 3.4: Benchmarking ANZSCO occupation measures 199
Appendix 3.5: Further evidence from sample attrition analyses 202

Appendix 3.6: ANZSCO required education ranges of jobs 204
Appendix 4.1: Human capital and job requirements in Australia 205
Appendix 4.2: Over-education by population sub-groups 208
Appendix 5.1: Definition of variables and descriptive statistics 213
Appendix 5.2: Full results from regression estimators 219
Appendix 5.3: Details for cross-sectional matching estimators 224
Appendix 5.4: Full results from combined matching and regression estimators 231
Appendix 6.1: Further descriptive statistics on persistence 234
Appendix 6.2: Wage penalty by duration over-educated 237
Appendix 6.3: Details for dynamic panel probit model estimates 241
Appendix 6.4: Prior over-education and wages of well-matched individuals 252
Appendix 7.1: Distributions of overall job satisfaction levels and intentions to quit 254
Appendix 7.2: Estimates of voluntary over-education 255
Appendix 7.3: Definition of variables and descriptive statistics 256
Appendix 7.4: Robustness analyses for differences in job attributes 258
Appendix 7.5: Descriptive statistics on persistence and voluntary over-education 260

List of tables

Table 3.1: Number of cases of incomplete information in highest education level variables—All individuals (N) 51

Table 3.2: Number of changes in highest education level by reported completion of study—Individuals who respond in consecutive waves (N) 52

Table 3.3: Number of cases of incomplete information in ANZSCO occupation variables—All employed individuals (N) 55

Table 3.4: Number of changes in occupation, based on various methods of identification—Individuals employed in consecutive waves (N) 56

Table 3.5: Number of revisions to ANZSCO occupation variables—All employed individuals (N) 58

Table 3.6: Number of 'definite' changes in occupation—Individuals employed in consecutive waves (N) 59

Table 3.7: Number of cases of incomplete information in revised ANZSCO occupation variables—All employed individuals (N) 60

Table 3.8: Probit model estimates for probability of non-response at *t* given individual responded at *t-1*—Attrition from two-year panels across W1–W8 63

Table 3.9: ANZSCO skill levels and corresponding formal education levels and years of relevant experience 66

Table 4.1: Incidence of over-education by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 72

Table 4.2: Incidence of over-education (using method that does not account for relevant work experience) by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 72

Table 4.3: Previous estimates of over-education in Australia—Employed individuals (%) 74

Table 5.1: Average wages by over-education status and gender 90

Table 5.2: Variation in wages and over-education status by gender 91

Table 5.3: Over-education wage penalty—ORU earnings functions, pooled OLS estimates 94

Table 5.4: Over-education wage penalty—ORU earnings functions, fixed effects and first-differences estimates 96

Table 5.5: Over-education wage penalty—Cross-sectional matching estimates 100

Table 5.6: Over-education wage penalty—ORU earnings functions, combined matching and pooled OLS estimates 101

Table 5.7: Over-education wage penalty—Difference-in-differences matching estimates 104

Table 5.8: Summary—Over-education wage penalty estimates (%) 106

Table 6.1: Over-education transition rates over one-year intervals by gender 113

Table 6.2: Cross-sectional incidences of over-education by total number of periods observed over-educated, gender and year—Individuals in balanced panel sample (%) 115

Table 6.3: Duration over-educated (at time *t*, not total) by gender and year—Individuals over-educated at *t* (%) 116

Table 6.4: Average wages by duration over-educated and gender 119

Table 6.5: Over-education wage penalty by duration over-educated—ORU earnings functions, fixed effects estimates 121

Table 6.6: Dynamic random effects probit model estimates for probability over-educated 131

Table 6.7: Estimates of state dependence in over-education 133

Table 6.8: Average wages by prior over-education status and gender—Well-matched individuals 137

Table 6.9: Wage effects of prior over-education among well-matched individuals—Difference-in-differences matching estimates 138

Table 7.1: Incidence of voluntary and involuntary over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 148

Table 7.2: Estimated differences in job attributes between over-educated and well-matched individuals 157

Table 7.3: Estimated differences in job attributes between over-educated and well-matched individuals—Using alternative thresholds to identify voluntary over-education 161

Table 7.4: Estimated differences in job attributes between over-educated and well-matched individuals—Analyses by gender 163

Table 7.5: Persistent over-education by voluntariness (at time *t*) and gender (%) 164

Table A3.1.1: ANZSCO—Classification structure 194

Table A3.1.2: ANZSCO—Major and Sub-major groups 195

Table A3.2.1: Cases where highest education level of individual decreases over time 196

Table A3.2.2: Alterations to highest education level variables 196

Table A3.3.1: Benchmarking highest education levels, HILDA vs. ABS Survey of Education and Training (SET)—All individuals aged 15–64 years [(a)] (%) 197

Table A3.3.2: Benchmarking highest education levels, HILDA vs. ABS Survey of Education and Work (SEW)—All individuals aged 15–64 years (%) 198

Table A3.4.1: Benchmarking ANZSCO occupations at 1-digit level, HILDA vs. ABS Labour Force Survey (LFS)—All employed individuals aged 15 years and over (%) 199

Table A3.4.2: Benchmarking ANZSCO occupations at 2-digit level, HILDA vs. ABS Census of Population and Housing—All employed individuals aged 15 years and over (%) 200

Table A3.5.1: Probit model estimates for probability of non-response in any subsequent wave given individual responded in W1—Attrition from balanced panel W1–W8 202

Table A3.5.2: Probit model estimates for probability of non-response at *t+1* given individual responded at *t* and *t-1*—Attrition from three-year panels across W1–W8 203

Table A3.6.1: ANZSCO required education ranges of jobs (and years of relevant experience to substitute for formal education) 204

Table A4.1.1: Highest education level by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 205

Table A4.1.2: Labour market experience by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 206

Table A4.1.3: Required education level of job (and years of relevant experience to substitute for formal education) by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 207

Table A4.2.1: Incidence of over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 208

Table A4.2.2: Incidence of over-education by demographic characteristics and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 209

Table A4.2.3: Incidence of over-education by human capital measures and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 210

Table A4.2.4: Incidence of over-education by job characteristics and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 211

Table A5.1.1: Definition of dependent and independent variables used in analyses 213

Table A5.1.2: Descriptive statistics for dependent and independent variables by over-education status and gender 216

Table A5.2.1: ORU earnings functions—Pooled OLS estimates 219

Table A5.2.2: ORU earnings functions—Fixed effects and first-differences estimates 222

Table A5.3.1: Logit models for probability over-educated—Propensity score model results 224

Table A5.3.2: Balancing tests results—Differences in mean characteristics of over-educated and well-matched individuals pre- and post-matching (using nearest neighbour and kernel matching weights) 227

Table A5.3.3: Samples pre- and post-matching by highest education level and gender 229

Table A5.3.4: Distribution of propensity scores in post-matching samples, and reliance on well-matched individuals to identify nearest neighbour matching estimates across distribution 230

Table A5.4.1: ORU earnings functions—Combined matching and pooled OLS estimates 231

Table A6.1.1: Transitions between mismatched, well-matched and not employed states over one-year intervals by year (%) 234

Table A6.1.2: Sources of over-education entries and exits over one-year intervals by year (%) 235

Table A6.1.3: Total number of periods observed over-educated by gender—Individuals in balanced panel sample (%) 236

Table A6.2.1: Variation in wages and duration over-educated by gender 237

Table A6.2.2: ORU earnings functions with duration over-educated—Fixed effects estimates 238

Table A6.2.3: Over-education wage penalty by duration over-educated—ORU earnings functions, pooled OLS estimates 240

Table A6.3.1: Dynamic random effects probit model estimates for probability over-educated—Males 241

Table A6.3.2: Dynamic random effects probit model estimates for probability over-educated—Females 243

Table A6.3.3: Probit models for probability over-educated in initial period—Initial conditions models results for Heckman and Orme estimators 245

Table A6.3.4: Dynamic random effects probit model estimates for probability over-educated—Orme and Wooldridge estimators using balanced panel sample 247

Table A6.3.5: Dynamic random effects probit model estimates for probability over-educated—Orme and Wooldridge estimators with extended specifications..... 249

Table A6.3.6: Estimates of state dependence in over-education—Robustness analyses..... 251

Table A6.4.1: Wage effects of prior over-education among well-matched individuals—ORU earnings functions, pooled OLS estimates 252

Table A6.4.2: Wage effects of prior over-education among well-matched individuals—ORU earnings functions, fixed effects estimates 253

Table A7.2.1: Incidence of voluntary over-education by gender and year using various thresholds for 'highly satisfied' and 'highly unlikely to quit'—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%) 255

Table A7.3.1: Definition of dependent variables (job attributes) used in analyses..... 256

Table A7.3.2: Descriptive statistics for dependent variables by over-education status 257

Table A7.4.1: Variation in voluntary and involuntary over-education identifiers..... 258

Table A7.4.2: Estimated differences in job attributes between over-educated and well-matched individuals—Sensitivity analyses using alternative estimators..... 259

Table A7.5.1: Persistent over-education (using duration over-educated) by voluntariness (at time *t*) and gender (%)..... 260

List of figures

Figure 4.1: Incidence of over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed 73

Figure 5.1: Distribution of propensity scores by over-education status and gender 99

Figure A5.3.1: Distribution of propensity scores by over-education status and gender 229

Figure A6.1.1: Survival functions—Proportion of individuals who remain over-educated by number of years since first observed over-educated 236

Figure A7.1.1: Distribution of overall job satisfaction levels—Employed individuals aged 15–64 years, excluding full-time students and self-employed 254

Figure A7.1.2: Distribution of intentions to quit in next 12 months—Employed individuals aged 15–64 years, excluding full-time students and self-employed 254

List of abbreviations

ABS	Australian Bureau of Statistics
ACT	Australian Capital Territory
ANZSCO	Australian and New Zealand Standard Classification of Occupations
APE	Average partial effects
AQF	Australian Qualifications Framework
ASCED	Australian Standard Classification of Education
ASCO	Australian Standard Classification of Occupations
ATSI	Aboriginal or Torres Strait Islander
ATT	Average treatment effect on the treated
CIA	Conditional independence assumption
COAG	Council of Australian Governments
Coeff.	Coefficients
CSA	Common support assumption
DID	Difference-in-differences
ESB	English speaking background
HILDA	Household, Income and Labour Dynamics in Australia
JA	Job analyst
N	Number of observations (or sample size)
NCVER	National Centre for Vocational Education Research
NESB	Non-English speaking background
NSW	New South Wales
NT	Northern Territory
OLS	Ordinary least squares
RM	Realised matches
SE	Standard errors
UK	United Kingdom
US	United States of America
VET	Vocational Education and Training
WA	Worker self-assessed

Chapter 1

Introduction

Modern societies devote considerable resources to the education of individuals.[1] The principal factors are government funding and the years individuals spend undertaking education. In Australia, for example, federal and state governments allocated 16 per cent of their total expenditure (approximately $71 billion) in 2009-10 to funding education (ABS, 2011). Meanwhile, legislation dictates individuals spend at least 11 years in school, though the vast majority today choose to spend additional years undertaking education.[2] With increasing numbers of individuals completing secondary and tertiary qualifications and governments continuing to emphasise the importance of educational attainment (as reflected in recent Council of Australian Governments (COAG) targets), these outlays on education have been, and will likely continue, rising over time.[3, 4]

Spending on education is typically justified by the argument that education is an investment for individuals and societies: its costs are offset by benefits that accrue over the lifetime of individuals. Education has many costs. Individuals face direct costs (e.g., tuition fees and textbooks), indirect costs (e.g., foregone earnings and leisure time) and psychological costs (e.g., effort or disutility associated with undertaking education), while costs for societies include government funding allocated to education, loss of output while individuals undertake education and income assistance paid to students (e.g., Austudy in Australia). Education also has many potential benefits. For individuals, these include increased earnings and (other) preferred job attributes (e.g., flexible work hours), a decreased chance of unemployment, improved health (e.g., the highly educated are more likely non-smokers, exercise regularly and eat a healthy diet) and the consumption (or lifestyle) benefits of being a student. Meanwhile, higher labour force participation rates, higher rates of innovation and technological progress, more productive workforces, reduced crime rates and greater

[1] Throughout this study, 'education' refers to formal education and training undertaken in schools, vocational colleges and universities that leads to certified qualifications (as defined in the Australian Qualifications Framework (AQF) and classified in the Australian Standard Classification of Education (ASCED)) (ABS, 2001; AQF Advisory Board, 2007).
[2] Minimum school leaving ages typically coincide with the completion of Year 9 or Year 10. For individuals who undertake further education, completing secondary education (i.e., Year 12) usually requires a total of 13 years in education, a Bachelor Degree 16 years and a Doctoral Degree around 20 years.
[3] In Australia, the proportion of the working-age population who completed secondary education increased from 52 per cent in 1984 to 73 per cent in 2012, while the proportion with a tertiary qualification rose from 36 to 59 per cent over the same period (ABS, 1994; ABS, 2012).
[4] Specifically, the COAG targets are: increase the proportion of 20–24 year olds who complete Year 12 to 90 per cent by 2015 (COAG, 2008b; COAG Reform Council, 2010); and, between 2009 and 2020, halve the proportion of 20–64 year olds without a qualification at the Certificate III level or above, and double the number of Diploma and Advanced Diploma qualifications completed (COAG, 2008a).

social cohesion are all potential benefits for societies that, ultimately, lead to higher economic growth rates and improved living standards (Card, 1999; Borland, Dawkins, Johnson and Williams, 2000; Hartog and Maassen van den Brink, 2007).

Since many of these benefits are realised via the employment of individuals, labour markets are critical for the accrual of benefits to investments in education. Labour market failures, therefore, can adversely affect the benefits that accrue to individuals and societies. In particular, their failure to facilitate the full utilisation of the human capital individuals derive from education—a phenomenon referred to as *over-education* (Duncan and Hoffman, 1981; Green, McIntosh and Vignoles, 1999; Hartog, 2000; McGuinness, 2006)—would diminish such benefits. The reasoning is as follows. Drawing on human capital theory, it is assumed each individual has human capital—mental and physical capabilities derived from education, work experience, on-the-job training, innate abilities and mental and physical health—that directly affects their labour productivity achieved in the workplace and, as a result, firm output.[5] Thus, for societies, over-education would result in an under-utilisation of the human capital available in the workforce, leading to productivity levels, economic growth rates and living standards below their potential. Meanwhile, for individuals, wages are determined by their stock of human capital, though it is assumed firms pay no wage premium for human capital that does not enhance an individual's labour productivity in a given job (i.e., wages reflect the value of individuals' labour productivity achieved in the workplace). Hence, individuals who have human capital from education that is under-utilised in their current job—deemed to be *over-educated*—would not receive the full benefits to their investments in education, at least in terms of increased earnings.[6] Over-education, therefore, is a relevant concern for government policy regarding education and for contemporary research into the returns to education (Card, 2001; Leuven and Oosterbeek, 2011). Its existence is the focus of this study.

Over-education occurs when jobs have *productivity ceilings*: thresholds beyond which output becomes unresponsive to the human capital possessed by individuals. That is, for each job, there is a particular level of human capital—referred to as the *required human capital level*—at which additional human capital ceases to increase output produced (or labour productivity achieved) by individuals. With a neoclassical view of labour markets, over-education may arise in the short run as labour markets adjust to changes in the supply or demand for labour. It may also arise if factors impede (perfect) competition in labour markets. In particular, when the conditions of perfect information

[5] Human capital theory is typically attributed to Mincer (1958), Schultz (1961) and Becker (1964); this study has drawn on such material, along with material from Rumberger (1987), Becker (1995), Preston (1997), Cahuc and Zylerberg (2004), Norris, Kelly and Giles (2005) and Ehrenberg and Smith (2006).

[6] 'Over-education' is related to, but quite distinct from, 'over-investment in education'. Over-investment concerns the comparison of costs and benefits: *private* costs and benefits identify over-investment among individuals; *social* costs and benefits identify over-investment in societies. Over-education, however, merely considers the benefits of education. Thus, over-education does not necessarily imply an over-investment in education.

and perfect factor mobility fail to hold, it can lead to situations in which firms do not, or cannot, customise jobs to fully utilise individuals' human capital and, at the same time, situations in which individuals are willing to accept jobs for which they are over-educated. Over-education resulting from labour market imperfections can be a persistent (or long run) phenomenon for individuals and societies.

Assuming levels of education (or qualifications) can be used to quantify the required human capital levels of jobs, instances of over-education can be identified: an individual is *over-educated* if their education (or highest qualification) level exceeds that considered necessary to perform their job. Alternatively, an individual whose education matches the required level is deemed *well-matched* and an individual with less education than required is *under-educated*. An example of over-education is a university graduate working as a cleaner: he or she undoubtedly has human capital from their degree that is not being utilised in their current job and, as a result, is likely to earn less than if instead employed in a professional-level job for which qualified.[7] Moreover, they could have obtained and performed the job with the completion of less education (or a lower level qualification), which presumably would have been less costly. Not all instances of over-education, however, are necessarily so extreme. For example, some may involve individuals with vocational qualifications in jobs requiring only secondary education.

This study investigates the existence of over-education in Australian labour markets. It seeks to determine whether there are labour market failures—inefficiencies in the matching of individuals and jobs—that lead to an under-utilisation of the human capital individuals derive from education. Three key questions are considered. Is there evidence of over-education in Australian labour markets? Is over-education merely a by-product of adjustment processes in dynamic and well-functioning labour markets (i.e., short-term disequilibria that have no enduring effects)? And, based on a broad assessment of the benefits of education, rather than focusing solely on earnings, is being over-educated actually the preferred outcome for some individuals (i.e., not the result of labour market failure)? These questions, and their implications, are further discussed below, but first the scope of this investigation is defined.

The study uses individual-level data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey—longitudinal data from a sample representative of the Australian population—for the 2001 to 2008 period to examine the individual (or private) benefits of education investments. Private costs are not examined and, as a result, the study does not attempt to identify over-investment in education among individuals. Similarly, the social costs and benefits of education

[7] Hence, the benefits to their investment in education are diminished. Notice, however, it is not argued that over-educated individuals necessarily receive zero benefits, because without their education (or highest qualification) they may have been unable to obtain and perform the job. But, zero benefits are possible. For example, if their second-highest qualification would have been sufficient for the job, then they receive no (work-related) benefit to their highest qualification.

are not examined and, therefore, the study does not consider over-investment at the societal level (e.g., whether social benefits are sufficient to justify the government funding allocated to education). In addition, given the focus on human capital derived from education, this study is not an assessment of the quality of individual-job matches with regards to all human capital. Limitations also arise from the empirical identification of over-education. Specifically, the study does not consider the potential for individuals' innate abilities, quality of educational institution attended and qualification vintage (or year completed) to affect the human capital derived from education.[8] Also, given each individual's highest qualification is used to identify over-education, the study does not consider whether having multiple qualifications renders an individual over-educated. And, due to data limitations, the quality of individual-job matches with regards to the subject matter (or field of study) of individuals' qualifications is not considered.

This study is structured as follows. Chapter 2 presents a synopsis of previous over-education research. It covers the theoretical foundation for over-education, the methods used to empirically identify it and empirical evidence relevant to the research questions of this study. Chapter 3 defines the research questions and describes the research method used to empirically test them. It describes the data used and its suitability for examining over-education in Australia, explains the method used to identify over-education and discusses the sample restrictions—ultimately, the sample examined is: employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed. Chapter 4 presents the first results. Specifically, it presents estimates of the incidence of over-education in Australian labour markets during the 2001–2008 period. Results indicate that approximately 20 per cent of working-age employees are over-educated in each year. The three subsequent chapters address the research questions of the study and, therefore, represent the main contributions to the literature.

Chapter 5 examines the key assertion regarding over-education: that individuals identified as over-educated have human capital from education that is under-utilised in their current job. Determining whether empirical evidence supports this assertion is critical for establishing evidence of over-education in Australian labour markets. If evidence supports it, instances of over-education represent a source of potential gains for the economy: having these individuals instead well-matched may lead to higher wages for the individuals and then higher productivity levels, economic growth rates and living standards. But, if there is no evidence to support it, over-education may be merely an artificial phenomenon identified in data—a statistical artefact (McGuinness, 2006)—which results from assumptions made regarding labour markets and the accumulation of human capital from education (i.e., assuming jobs have productivity ceilings and that all individuals acquire the same

[8] This is the key limitation in all over-education research. It is further discussed in Section 2.3.

human capital from a given qualification). Thus, individuals deemed over-educated are actually employed in jobs fully utilising their human capital from education and do not represent a source of potential gains.

In deriving empirical evidence, the focus is individuals' labour productivity achieved in the workplace, where evidence of reduced labour productivity is considered indicative of under-utilised human capital. It is assumed wages reflect the labour productivity achieved, and hence the relationship between over-education and wages is examined. Finding over-education has a negative, statistically significant effect—being over-educated *causes* an individual's wage to be lower than if they were instead well-matched (i.e., an *over-education wage penalty* exists)—therefore validates the assertion. Estimating such *causal effects*, however, is not straightforward.[9] The concern is that unobserved individual heterogeneity (or non-random selection into over-education) may lead to biased estimates. In response, a series of econometric techniques, some of which are designed to control for unobserved individual heterogeneity, are used to derive estimates. The robustness of these estimates is then examined, with the aim of establishing over-education wage penalty estimates that can be interpreted as causal effects. Estimates are derived using pooled OLS, fixed effects, first-differences, cross-sectional (or propensity score) matching, combined matching and regression, and difference-in-differences (or longitudinal) matching estimators. Chapter 5 is one of only a few studies to empirically test this assertion. It also adds to the few studies acknowledging the need to account for unobserved individual heterogeneity when estimating the over-education wage penalty, and its use of detailed data and sophisticated econometric techniques means more reliable estimates are contributed to previously unclear evidence. Ultimately, results indicate that individuals identified as over-educated do indeed have human capital from education that is under-utilised in their current job.

Chapter 6 considers the possibility over-education is a result of the inherently dynamic nature of labour markets. Modern labour markets are complex and continuously evolving: there are always individuals investing in human capital, moving in and out of the labour force and moving between jobs (i.e., labour supply is constantly changing), and firms are always creating, adjusting and discontinuing jobs (i.e., labour demand is constantly changing). This may result in individual-job mismatches that lead to over-education. But, with time, it may also lead to their resolution as either individuals, in seeking to maximise their utility (or wages), move to jobs that fully utilise their human capital, or firms, in seeking to maximise their profits, adjust jobs to fully utilise the available human capital. Hence, instances of over-education may be short-term disequilibria. If the time spent over-educated has no enduring effects, over-education could then be regarded as merely a by-product of

[9] *Causal effects* are defined using the potential outcomes framework and estimated using parametric and semi-parametric techniques, where conditional independence is critical for identifying causal parameters. This is consistent with the approach of leading researchers in applied economics (see, for example, Heckman (2008), Angrist and Pischke (2009) and Imbens and Wooldridge (2009)). Estimation of causal effects is further discussed in Section 5.3.

adjustment processes in dynamic and well-functioning labour markets. In which case, government policy interventions, if deemed necessary, should be aimed at increasing the speed of such adjustments to minimise the costs of over-education.[10] If, however, over-education is persistent for individuals or has enduring effects, then it represents more serious labour market failures, and government policy interventions would be necessary to prevent and resolve it. The most likely enduring effect, which is considered in this study, is that time spent over-educated could lead to human capital depreciation. Such a link would mean the costs of over-education are greater than first thought as they not only arise while individuals are over-educated, but also in future employment where individuals appear well-matched (i.e., their labour productivity will be lower than if they had not previously been over-educated).

The chapter performs two sets of analyses. The first considers whether over-education is merely short-term labour market disequilibria. In particular, it examines individuals' transitions to and from over-education and their durations over-educated, and then estimates the over-education wage penalty by individuals' duration over-educated. The fixed effects estimator is used in an attempt to derive estimates that can be interpreted as causal effects. Finding individuals are only temporarily identified as over-educated or the wage penalty disappears over time (for those who remain identified as over-educated) is assumed to be evidence that instances of over-education are short-term disequilibria. The second set of analyses considers whether over-education has enduring effects, and specifically, the possibility it leads to human capital depreciation. Two empirical tests are performed. One examines whether being over-educated in the past increases the likelihood of being over-educated again in the future (i.e., tests for state dependence in over-education). Dynamic panel probit models (and the Heckman (1981c), Orme (1997) and Wooldridge (2005) estimators) are used to derive causal effects estimates. The other empirical test examines the relationship between prior over-education and the wages of currently well-matched individuals. The difference-in-differences matching estimator is used to estimate the effect prior over-education has on individuals' wages. Finding state dependence in over-education or prior over-education reduces the wages of well-matched individuals is assumed to be evidence that over-education leads to human capital depreciation and, therefore, has enduring effects. Chapter 6 addresses an important weakness in the over-education literature, which is that studies are typically based on static analyses of labour markets. As a result, it adds to limited empirical evidence on the persistence of over-education, particularly the information on individuals' durations over-educated, and it is the first study to examine dynamics of the over-education wage penalty. Also, it is the first to test for state dependence in over-education,

[10] As outlined above, the costs of over-education include diminished productivity levels, economic growth rates and living standards for societies and reduced earnings (and possibly reduced utility levels) for individuals.

with the evidence contributing to limited research on the potential link between over-education and human capital depreciation.

Chapter 7 considers the possibility not all instances of over-education represent labour market failures. Given many attributes of jobs (e.g., work hours, job security and required effort), and not just the wage, can affect individuals' utility levels, it is possible some over-educated individuals have obtained jobs that maximise their (expected) utility levels and, therefore, achieved their preferred outcome. Such individuals, referred to as *voluntarily over-educated* in this study, would be trading wages for non-pecuniary benefits of employment or improved working conditions and, as a result, accepting jobs for which they are over-educated. Since *voluntary over-education* arises due to individuals' preferences regarding job attributes, it would not be considered the result of labour market failures, at least from the perspective of individuals.

It is first necessary to develop a method by which instances of voluntary over-education, if they exist, can be identified. The chapter proposes a distinction based on individuals' job satisfaction levels and their desire for a new job (or intentions to quit current job): it is assumed over-educated individuals who are highly satisfied with their job and highly unlikely to quit are *voluntarily over-educated*, while the remainder are *involuntarily over-educated*. This method and the resultant estimates must then be validated. If valid, there should be evidence of trade-offs between wages and other job attributes among the individuals identified as voluntarily over-educated, and hence the relationship between voluntary over-education and job attributes is examined. Finding that, compared to well-matched individuals, the voluntarily over-educated experience wage penalties and improvements in other job attributes (and the involuntarily over-educated incur wage penalties without such improvements) is assumed to be evidence that some individuals are indeed voluntarily over-educated. Empirical evidence is derived from the estimation of a series of econometric models with job attributes as dependent variables. Specifically, linear fixed effects, fixed effects ordered logit and fixed effects probit estimators are used to derive estimates. Each controls for (observable and unobservable) individual heterogeneity to ensure the estimated differences in job attributes are not confounded by other factors (i.e., reflect only the differences that arise from being voluntarily (and involuntarily) over-educated rather than well-matched). Chapter 7 makes several contributions to the literature. Most importantly, it responds to the limitation that previous over-education research has focused on increased earnings as the only (private) benefit to education investments, and thereby overlooked the role individuals' preferences may play in over-education. As a result, it is the first study to attempt to estimate the incidence of voluntary over-education, and the first to assess the relationship between such voluntary over-education and job attributes.

Chapter 8 concludes the study. It discusses the conclusions drawn and highlights suitable avenues for future research.

Chapter 2

Over-education research

2.1 Introduction

The origin of economics research into over-education is typically considered to be Richard Freeman's (1976) book *The Overeducated American*, in which evidence of declining returns to college education (i.e., decreased wage differentials between graduates and non-graduates) was used to support the argument that growth in the supply of college graduates had far exceeded demand in the early 1970s in the US.[1] Ultimately, Freeman (1976) concluded it was an era of *over-education*, but did not estimate the incidence of such over-education. The first study to do so was Duncan and Hoffman (1981). It used information collected in the Panel Study of Income Dynamics in the US in 1976 to derive a measure for the required education level of each individual's current job, and then compared it to actual education levels to classify each individual-job match and, thereby, estimate the incidence of over-education.[2] Roughly 40 per cent of the US workforce was deemed over-educated. In addition, this over-education measure was incorporated into a Mincer (1974) earnings function and it was found that returns to excess education were statistically significant and positive, but smaller in magnitude than returns to required education. As a result, Duncan and Hoffman (1981) concluded that over-educated individuals may be failing to earn the full returns to their completed education. This finding was, almost certainly, the catalyst for most subsequent research in this area.[3]

In the years since Freeman (1976) and Duncan and Hoffman (1981), over-education has been the subject of considerable economics research involving many different datasets and countries.[4] This literature has predominantly focused on three topics: theoretical frameworks to explain the existence of over-education in labour markets; methods to empirically identify over-education and resultant estimates of its incidence; and estimation of the effect over-education has on

[1] Ivar Berg's (1971) book *Education and Jobs: The Great Training Robbery* examined a similar topic, but it is sociological rather than economic in its approach. The study of over-education can also be linked to research into skills mismatch (i.e., the matching of heterogeneous individuals to heterogeneous jobs), and this dates back to Tinbergen (1956).
[2] Specifically, Duncan and Hoffman (1981) used responses to the survey question '*How much formal education is required to get a job like yours?*' to identify instances of over-education.
[3] Given most subsequent research also emulated the analyses (i.e., used measures for required education levels to estimate over-education, then estimated the returns to required and excess education using extended Mincer earnings functions), Duncan and Hoffman (1981) more aptly represents the pioneer study for economics research into over-education.
[4] Throughout this study, it is this body of economics research that is referred to as the *over-education literature*.

individuals' wages—the *over-education wage penalty*.[5] Each topic is discussed here. This chapter also reviews previous over-education research that is relevant to this study. Specifically, the data, analyses, results and limitations of studies examining similar issues are discussed. This chapter is not intended to be an exhaustive review of the over-education literature. Instead, it discusses issues important to the study of over-education and empirical evidence relevant to the research questions of this study. More extensive reviews and evaluations can be found in McGuinness (2006) and Leuven and Oosterbeek (2011).[6]

The chapter proceeds as follows. Section 2.2 outlines the theoretical foundation for over-education. Section 2.3 considers the empirical identification of over-education. It discusses the three methods used in the literature and argues there is one that should be preferred; that preferred method is used in this study. Section 2.4 discusses estimation of the over-education wage penalty. Section 2.5 considers research into the dynamics of over-education. In particular, evidence regarding the persistence of over-education for individuals and the possible link between over-education and human capital depreciation. Section 2.6 discusses research concerning over-education and individuals' preferences (i.e., the possibility of *voluntary over-education*). Section 2.7 provides an overview of the research examining over-education in Australian labour markets. Section 2.8 concludes the chapter by highlighting key limitations of the over-education literature; it aims to identify the areas in which important contributions can be made.

2.2 Theoretical foundation for over-education

In the over-education literature, no single theoretical framework has been established to explain the existence of over-education. Instead, studies have typically been based on one (or more) of the following: *human capital theory*; *assignment models*; *job signalling model*; or, *job competition model*. This lack of consensus is a key weakness in the literature because it means there is no clear explanation for why over-education may occur, which then makes it difficult to propose government policy interventions

[5] Other topics have been considered. The most widely examined include: the identification of individual characteristics associated with being over-educated (via the estimation of probability models) (see Sloane (2003) and Leuven and Oosterbeek (2011) for overviews of the resultant evidence); the relationship between over-education and individuals' job satisfaction levels (see, for example, Tsang, Rumberger and Levin (1991), Fleming and Kler (2008) and Green and Zhu (2010)); and the relationship over-education has with *skills mismatches* (see, for example, Allen and van der Velden (2001) and Green and McIntosh (2007)) and *over-skilling* (see, for example, Mavromaras, McGuinness and Fok (2009) and McGuinness and Wooden (2009)). The analysis of skills mismatches and over-skilling is discussed in Section 2.3.

[6] There are several other notable studies and books. First, Green et al. (1999) provide insights into the identification and interpretation of over-education. Second, 'The Economics of Over- and Underschooling' special issue of *Economics of Education Review* (Volume 19, Issue 2) contains studies that are representative of the analyses and findings in the literature (see Oosterbeek (2000) for an overview of this special issue and Hartog (2000) for an appraisal of the literature). Third, the Borghans and De Grip (2000) edited book *The Overeducated Worker? The Economics of Skill Utilisation* considers theoretical foundations for over-education and whether it leads to the less educated being crowded out of labour markets. Fourth, the Büchel, De Grip and Mertens (2003) edited book *Overeducation in Europe: Current Issues in Theory and Policy* provides a further overview of empirical methods and findings in the literature, particularly the Sloane (2003) chapter.

to prevent and resolve any such over-education (Hartog, 2000; McGuinness, 2006; Leuven and Oosterbeek, 2011). The weakness arises because modern economics research, more generally, uses numerous theoretical frameworks to study labour market phenomena, and it is difficult to distinguish between them (i.e., empirical predictions that can clearly indicate which theory best explains the functioning of labour markets are scarce). Since over-education may occur in many of these theories, this general uncertainty regarding a single (unified) theoretical framework for labour markets leads to similar uncertainty in the over-education literature.[7] Given the empirical focus of this study, the theoretical foundation for over-education is only briefly discussed here.[8]

As stated in Chapter 1, an individual who has human capital from education that is under-utilised in their current job is deemed *over-educated*. For such over-education to occur, it must be assumed that jobs have *productivity ceilings*: thresholds beyond which output becomes unresponsive to the human capital possessed by individuals. That is, for each job, there is a particular level of human capital—referred to as the *required human capital level*—at which additional human capital ceases to increase the output produced (or labour productivity achieved) by individuals. It is the level required to *maximise* output (or labour productivity) achieved in the job. An alternative interpretation is that it is the level required to competently perform the job (i.e., the *minimum* level necessary to be productive in the job); meaning individuals with less human capital are unable to perform the job.[9] The *minimum* interpretation is the one most prevalent in the over-education literature, and it is used because studies also seek to identify under-educated individuals. In fact, in studies that categorise employed individuals as over-educated, under-educated and well-matched, both interpretations are used simultaneously. That is, the required human capital levels of jobs are interpreted as both the *minimum* level required to perform the job (as individuals with human capital from education below this level are deemed under-educated) and the level required to *maximise* output (as individuals with human capital from education exceeding this level are deemed over-educated).[10] For the following

---

[7] Empirical evidence on the dynamics of over-education (e.g., individuals' entries, exits and durations spent over-educated) has the potential to help determine which theory best explains how over-education may arise. But, despite the relatively large amount of over-education research, little is known about such dynamics. Nevertheless, several studies have sought to determine the most appropriate theoretical model for over-education (see, for example, Sloane, Battu and Seaman (1999), Groot and Hoek (2000), Allen and van der Velden (2001), Büchel and Mertens (2004), Groeneveld and Hartog (2004), Linsley (2005), Di Pietro and Urwin (2006) and Green and McIntosh (2007)). These studies typically focused on theoretical predictions for the wage effects of education and then estimated wage equations to test such predictions. The resultant empirical evidence has not established any one theory as being preferred to all others.

[8] That is, since this is an empirical thesis (motivated by the limitations of existing empirical studies), there is only a brief (intuitive) discussion of the theoretical foundation for over-education. No formal theoretical model is presented.

[9] Since jobs can have tasks that require varying levels of skill (e.g., the role of university professor involves teaching and research tasks that require human capital derived from a Doctoral Degree, but also administrative tasks that require much less human capital), then this *minimum* interpretation implicitly assumes that the most difficult tasks determine the required human capital levels of jobs.

[10] In this study, under-educated individuals are identified (principally to exclude them from the main analyses) and so it too uses both interpretations of required human capital levels. However, unlike most other studies, the method used to empirically identify over-education actually reflects this as required human capital *ranges* (i.e., a minimum and a maximum) are defined for each job. These issues are further discussed in Sections 2.3 and 3.5.

discussion of the theoretical foundation for over-education, it is sufficient to interpret the required human capital levels of jobs as the level at which output (or labour productivity) is maximised.

Assuming the existence of such rigid productivity ceilings, however, may be too simplistic. In reality, the relationship between human capital and output may be better characterised as a continuous production function with diminishing returns to human capital (i.e., output increases with greater human capital, but at a decreasing rate). This relationship—the shape of the production function—is also likely to vary across jobs. For some jobs, the marginal returns to human capital may diminish only slightly—the production function is essentially a straight line—and, therefore, no productivity ceilings exist. High-skilled jobs (i.e., senior management and professional jobs) and jobs of self-employed individuals are possible examples.[11] For other jobs, the marginal returns to human capital may diminish significantly and at a certain point become relatively minor (or effectively zero)—the production function is essentially a flat line from a certain point onwards—which is analogous to productivity ceilings existing.[12]

A further issue concerns the relevance of human capital: for each job, it may be only human capital relevant to the job that increases output, whereby relevance is with respect to the field or subject matter to which human capital and jobs can be characterised (e.g., physics, civil engineering, political science, nursing and accounting). For instance, the human capital derived from a nuclear physics qualification is unlikely to increase the output produced by an individual working in an aged care nursing job. This means it is possible for individuals to encounter productivity ceilings because some of their human capital is irrelevant for the particular job. Hence, the required human capital levels of jobs are with respect to *relevant* human capital only and should be interpreted accordingly (i.e., they represent the level of *relevant* human capital required to maximise the output produced in the job).[13] Despite these potential complications, the key point here is that over-education can only occur when jobs have productivity ceilings that can prevent individuals from fully utilising their human capital in performing the job.

---

[11] In both cases, it is likely jobs are flexible and individuals have autonomy in designing the job. Hence, the jobs can be tailored to match the individuals' human capital and, as a result, additional human capital always leads to increased output (or labour productivity). This study accounts for the likelihood that such jobs do not have productivity ceilings; specifically, required human capital levels of management and professional jobs are defined such that individuals employed in them cannot be deemed over-educated, while self-employed individuals are omitted from the analyses. The sample restrictions and empirical identification of over-education are further discussed in Sections 3.3 and 3.5.

[12] With this more nuanced view of the relationship between human capital and output, the required human capital levels of jobs can be interpreted as a form of *optimal* human capital level in terms of balancing the costs and benefits of human capital investments (Verhaest and Omey, 2006b). Alternatively, they can be interpreted as the point at which only a fraction of additional human capital increases the output (or labour productivity) achieved in the job (i.e., the point from which additional human capital is not *fully* utilised in performing the job).

[13] Due to data limitations, which are further discussed in Section 3.3, this study does not consider quality of individual-job matches with respect to subject matter (or, as defined in the ASCED, the *field of education* of individuals' qualifications) and the possibility mismatches may result in an under-utilisation of human capital. As a result, this study may under-estimate the extent to which human capital from education is under-utilised in Australian labour markets.

With a neoclassical view of labour markets, over-education may arise in the short run as labour markets adjust to changes in the supply or demand for labour.[14] The prime example is an increase in the supply of individuals with a university degree: assuming jobs (or labour demand) remain unchanged in the short run, the excess supply of university graduates leads to a decline in their wage relative to less educated individuals.[15] Firms observe the relative wage reduction and consequently substitute some of their less educated workers for graduates, as they presumably have a greater productive capacity. Since jobs are unchanged, the graduates employed in jobs previously performed by less educated individuals are therefore over-educated.[16] These graduates accept over-educated jobs because, in the short run at least, based on constrained choice sets—the set of obtainable jobs and the choice to not work—it maximises their (expected) utility. But, since jobs fully utilising their human capital would presumably result in higher utility, it can be considered a form of second-best solution (i.e., merely the best job they can obtain in the short run, and it is preferred to non-employment). A further possibility is that such jobs are a pathway to future career advances. This argument is put forward in *career mobility theory*, as developed by Sicherman and Galor (1990) and Sicherman (1991), whereby education not only has the direct effect of increasing individuals' earnings, but also an indirect effect of increasing their likelihood of future career advancements. Since individuals seek to maximise (expected) returns to education over their entire careers, it may be optimal to accept jobs with a relatively low direct return (i.e., wage) if the indirect return (i.e., chance of moving to jobs with higher wages) is greater than in other feasible jobs. Ultimately, over-education is only short-term disequilibria because labour markets eventually adjust to the change in (relative) supply of graduates. That is, either the firms, in seeking to maximise profits, adjust jobs to fully utilise the human capital of these graduates, or the graduates, in seeking to maximise utility (or wages), move to jobs that fully utilise their human capital. Over-education, therefore, arises as a by-product of short run adjustment processes in labour markets.[17]

Over-education may also arise if factors impede (perfect) competition in labour markets; specifically, when the conditions of perfect information and perfect factor mobility fail to hold. For

---

[14] In adopting a neoclassical view of labour markets, human capital theory is essentially being used to explain the existence of over-education. This is done because it is the approach (or labour market model) most often used in modern economics research to study wage determination and individuals' decisions to undertake education (Preston, 1997; McGuinness, 2006; Hartog and Maassen van den Brink, 2007). For a discussion of how over-education arises in alternative theoretical frameworks—assignment models, job signalling model and job competition model—see Appendix 2.1.

[15] In this situation, productivity ceilings exist because it is assumed jobs are fixed in the short run.

[16] Given the greater productive capacity of these graduates is not immediately utilised, as jobs are fixed, firms likely view their hiring as an investment: they are a source of future productivity gains (if firms choose to adjust the jobs), a pool from which to make future promotions and insurance against future labour shortages (Hersch, 1995; Büchel, 2002).

[17] Since it results from rational decisions by firms and individuals, career mobility theory asserts that over-education is an equilibrium outcome in dynamic and well-functioning labour markets: it is temporary for individuals—most likely occurring at the start of their careers—and requires no policy interventions to be prevented or resolved (Sicherman and Galor, 1990; Büchel and Mertens, 2004; Linsley, 2005). Such a claim, however, is dependent on over-education having no detrimental effects on individuals' subsequent employment outcomes. Recall, this claim is the focus of Chapter 6 in this study.

over-education to occur, these failures must lead to situations in which firms do not, or cannot, customise jobs to fully utilise individuals' human capital (i.e., productivity ceilings exist) and, at the same time, situations in which individuals are willing to accept jobs for which they are over-educated. Consider first the existence of productivity ceilings. Imperfect (or asymmetric) information regarding human capital can result in such productivity ceilings because the inability of firms to perfectly observe individuals' human capital may prevent them from (perfectly) adjusting or designing jobs to fully utilise the human capital of employees.[18] Imperfect factor mobility can also result in productivity ceilings. In particular, the existence of adjustment costs (e.g., costs associated with hiring and firing individuals and purchasing new production technologies) may mean the costs of customising jobs exceed the potential benefits (i.e., value of increased firm output). In such cases, firms would choose to not customise jobs. Additionally, labour market institutions, such as worker trade unions and collective bargaining agreements, may prevent firms from altering or specialising jobs. In these cases, firms cannot customise jobs. Productivity ceilings may also arise in jobs that have a fixed set of tasks to be performed and, as a result, a fixed wage. This is most likely the case in low-skilled jobs (e.g., cashiers, cleaners and shelf fillers), but such inflexibility may also arise in higher skilled jobs (e.g., safety inspectors, police officers and accounting clerks). In all of the above cases, since it is assumed firms pay no wage premium for unused human capital, the firms hire over-educated individuals because it is no more costly than hiring individuals whose education matches job requirements. And, as was the case with short-run over-education, firms may view the hiring of over-educated individuals as a form of investment.[19]

These labour market imperfections can also lead individuals to accept jobs for which they are over-educated. Imperfect (or asymmetric) information is an obvious cause: the inability of individuals to perfectly observe the human capital requirements of jobs, which is particularly likely prior to performing a given job, may result in them accepting over-educated jobs. Thus, individuals do not realise, at least initially, that they are over-educated for the job.[20] Imperfect factor mobility, specifically the existence of adjustment costs (e.g., costs associated with job search and geographical relocation), can also lead to the acceptance of over-educated jobs. For instance, individuals may be spatially constrained because relocation costs exceed the expected benefits (i.e., increased income from obtaining a job that fully utilises their human capital).[21] Hence, similar to the case of short-run

[18] This can lead to firms using education (or qualifications) as a signal for individuals' human capital levels, as in the job signalling model of Spence (1973). For a discussion of this model, and how over-education can arise in it, see Appendix 2.1.
[19] Of course, where there is imperfect information, firms may be unaware that they have hired over-educated individuals.
[20] In such situations, individuals must also be unable to perfectly observe the market-determined wage for their human capital because otherwise, assuming they seek to fully utilise their human capital (i.e., maximise their utility), they would consider any job with a wage below their 'market-value' as sub-optimal and, therefore, not accept it.
[21] The possibility spatial constraints lead to over-education was explored in Frank (1978) and McGoldrick and Robst (1996); in particular, it was argued that the job search of married women, especially those who are the secondary earner in a family, is geographically constrained (to where husband is employed), which can lead them to accept over-educated jobs.

over-education, individuals accept over-educated jobs because, given their (constrained) choices, it is the best job they can obtain.[22] A further possibility is that over-educated jobs are actually the preferred outcome for some individuals. Since many attributes of jobs can affect individuals' utility levels, such as the wage, work hours, job security and required effort, it is possible some over-educated individuals have obtained jobs that maximise their (expected) utility levels and, therefore, achieved their preferred outcome. Thus, given their preferences regarding job attributes, some individuals may trade wages for non-pecuniary benefits of employment or improved working conditions and, as a result, accept jobs for which they are over-educated.[23, 24] Ultimately, over-education resulting from labour market imperfections, and possibly individuals' preferences, will persist as long as factors impede (perfect) competition in labour markets. Over-education, therefore, can be a persistent (or long run) phenomenon for individuals and societies.

Regardless of how it arises and whether it is temporary or persistent, over-education would diminish the benefits to investments in education that accrue to individuals and societies over time. For societies, since individuals' human capital directly affects their labour productivity achieved in the workplace and firm output levels, over-education would result in an under-utilisation of the human capital available in the workforce, leading to productivity levels, economic growth rates and living standards below their potential. For individuals, since wages are determined by their stock of human capital and firms pay no wage premium for unused human capital, over-educated individuals would not receive the full benefits to their investments in education, at least in terms of increased earnings.[25] But, as discussed above, a broader assessment of the benefits of education, which recognises other job attributes as potential benefits, may reveal that this is actually the preferred outcome for some individuals. Nevertheless, one key implication remains: the extent to which individuals and societies benefit from investments in education is dependent on the quality of the individual-job matches, with respect to human capital utilisation, which are realised in labour markets.

---

[22] Labour market discrimination (e.g., on the basis of gender or ethnicity) can also constrain individuals' choice sets in employment decisions and, therefore, lead individuals to accept jobs for which they are over-educated.

[23] This is consistent with the theory of *compensating wage differentials* (or *equalising differences*) in labour markets (see, for example, Rosen (1986), Biddle and Zarkin (1988), Hwang, Reed and Hubbard (1992) and Bender (1998)). And there is empirical evidence to support it. In particular, it has been found that highly skilled individuals are more likely employed in jobs offering favourable working conditions; hence, they appear willing to *spend* some of their greater earnings capacity to obtain (their) preferred job attributes (Brown, 1980; Rosen, 1986).

[24] It is also similar to the argument proposed by career mobility theory. Importantly, however, career mobility theory argues over-educated jobs are a pathway to future career advances and so individuals should be in them only temporarily, but that is not necessarily the case here. The potential for such *voluntary over-education* is further discussed in Sections 2.6 and 3.2 and, recall, its existence is the focus of Chapter 7 in this study.

[25] In these circumstances, each individual's wage depends on the particular job they are employed in and the extent to which it utilises their human capital. Empirical evidence supports this (see, for example, Gruetter and Lalive (2009)).

2.3 Empirical identification of over-education

As previously discussed, comparing the human capital each individual derives from education with the required human capital level of their job determines whether individuals are over-educated. Empirical identification of such over-education is reliant on two key assumptions. The first is that levels of education (or qualifications) can be used to quantify the required human capital levels of jobs—thus, a *required education level* is defined for each job.[26] The second assumption is that individuals who complete the same level of education (or qualification) acquire the same level of human capital from it. Given these assumptions, over-education is empirically identified using information on the level of education completed by each individual and the required education level of their current job. In particular, an individual is deemed *over-educated* if their education (or highest qualification) level exceeds that considered necessary to perform their job.[27] Alternatively, an individual with less education than necessary is deemed *under-educated*.[28] Collectively these two states are often referred to as *educational mismatches*. Finally, an individual whose education matches the required education level of their job is deemed *well-matched*.

The empirical identification of over-education, as described above, leads to limitations in over-education research. First, given the two key assumptions, over-education research does not consider the potential for individuals' innate abilities, quality of educational institution attended and qualification vintage (or year completed) to affect the human capital derived from education (Halaby, 1994; Green et al., 1999; Chevalier, 2003; Van der Meer, 2006; Verhaest and Omey, 2006b). While such factors are not accounted for in the identification of over-education, they can, however, be considered when examining its incidence or controlled for when analysing its effects (e.g., in multivariate analyses estimating the over-education wage penalty). A further limitation, given the use

[26] Consistent with required human capital levels, these required education levels refer exclusively to jobs and the human capital necessary to perform their tasks. Hence, the characteristics of the individuals in each job are irrelevant and a single required education level is defined for each job.

[27] In the literature, over-education has been empirically identified with comparisons in terms of years of education (see, for example, Bauer (2002), Kler (2005), Korpi and Tåhlin (2009) and Tsai (2010)) and in terms of qualification levels (see, for example, Büchel and Mertens (2004), Frenette (2004), Green and McIntosh (2007) and McGuinness and Sloane (2011)). In this study, qualification levels are used because they more accurately capture the human capital derived from education and are the more relevant measure in modern labour markets. This issue is further discussed in Section 3.3.

[28] The relevance of such under-education, however, is limited. Since it is identified in reference to required human capital, it is typically interpreted as instances in which individuals have insufficient human capital to perform their job. But, such an interpretation is inaccurate. While it indeed identifies instances where individuals' human capital from education is insufficient to perform their job, this does not necessarily mean their entire stock of human capital is insufficient. This is because individuals can derive the human capital necessary to perform their job from work experience (and on-the-job training) rather than education. Thus, the method used to identify under-education overlooks the potential in modern labour markets to substitute work experience for formal education. To produce a more meaningful counterpart to over-education—under-education that indeed represents individuals with insufficient human capital to perform their job—it is, therefore, necessary to account for this potential substitution. To the author's knowledge, no previous studies have taken such an approach in identifying educational mismatches. The identification and interpretation of under-education in this study is further discussed in Section 3.5.

of each individual's highest qualification to identify over-education, is that no consideration is given to whether the completion of multiple qualifications may render an individual over-educated. Such qualifications could be at the same level or different levels.[29] Finally, over-education research does not consider whether the human capital individuals derive from education is entirely relevant for their job, in terms of subject matter (Halaby, 1994; Borghans and De Grip, 2000; Sloane, 2003).[30] Recall from Section 2.2, individual-job matches should be evaluated with regards to *relevant* human capital, but, since the empirical identification is based solely on levels of education, over-education studies do not consider such relevance. Each of these limitations may affect the validity of over-education research. In particular, the ramifications of the first limitation are the most serious: if the human capital derived from education varies by individuals' innate abilities, quality of educational institution attended or qualification vintage, then the method of empirical identification may be inadequate for identifying individuals with under-utilised human capital (which would then affect the validity of all subsequent over-education research).[31] The other limitations, however, may merely lead over-education research to under-estimate the extent to which there is human capital from education under-utilised in labour markets.[32, 33]

---

[29] For instance, an individual with two qualifications at the same level (e.g., Bachelor Degree in Business Management and Bachelor Degree in Nursing) whose current job only uses human capital from one of them (e.g., working as a Nurse) may be considered over-educated. Meanwhile, an individual with two qualifications at different levels (e.g., Advanced Diploma and Certificate IV) may also be considered over-educated if their current job does not use human capital from the lower qualification (and provided it was not a pre-requisite for the higher qualification).

[30] Subject matter refers to the field of study (or, as defined in the ASCED, the *field of education*) of individuals' qualifications (e.g., microbiology, computer science, sales and marketing, psychology and hairdressing).

[31] For example, if individuals derive less human capital from a degree completed at a low-ranked university (compared to a world-class university), then some of the individuals identified as over-educated (those who attended low-ranked universities) may not have under-utilised human capital. Instead, they may actually be well-matched.

[32] For a summary of the key definitions, assumptions and limitations in this study of over-education see Appendix 1.1.

[33] In the literature, many studies have assumed that all individuals with the same level of education have the same level of human capital (i.e., they are perfect substitutes for one another and represent a homogenous stock of potential workers) (Halaby, 1994; Green et al., 1999; Battu, Belfield and Sloane, 2000; Chevalier, 2003; McGuinness, 2006; Verhaest and Omey, 2006b; Green and McIntosh, 2007; Leuven and Oosterbeek, 2011). Given the definition (and interpretation) of over-education in this study, such an assumption—referred to as the *worker homogeneity assumption*—is unnecessarily restrictive. Instead, as stated above, the necessary assumption is that all individuals acquire the same level of human capital from a given level of education (or qualification). The worker homogeneity assumption is also unlikely to hold because, for example, increased access to education and greater numbers of individuals with a given qualification mean there is likely heterogeneity in innate ability levels of those with the same education and, therefore, heterogeneity in their level of human capital (Chevalier, 2003). The worker homogeneity assumption arose because standard practice in the literature was to include over-education measures in earnings functions, whereby, in order for resultant estimates to represent the causal effect of over-education on individuals' wages, a similar assumption was necessary (i.e., assumption of *conditional independence* or *no omitted variable bias*). Estimation of earnings functions and the necessary assumptions are discussed in Section 2.4. Given the limitations it imposed on subsequent research, the worker homogeneity assumption led to research examining the utilisation of human capital via estimates of *skill mismatches*, which were typically identified using survey questions on the extent to which individuals were utilising their abilities, knowledge and skills (not just from formal education) in performing their jobs. Some of these studies compared skill mismatch with over-education to empirically test whether the over-educated do indeed have under-utilised human capital (see, for example, Allen and van der Velden (2001), Chevalier and Lindley (2006), Di Pietro and Urwin (2006), Allen and de Weert (2007) and Green and McIntosh (2007)), while others formed a parallel stream of research examining the issue of *over-skilling* (see, for example, Mavromaras et al. (2009), McGuinness and Wooden (2009), Mavromaras, McGuinness, O'Leary, Sloane and Fok (2010b), McGuinness and Sloane (2011) and Mavromaras and McGuinness (2012)). The worker homogeneity assumption also led to the development of alternative methods for identifying over-education. The first, presented in Chevalier (2003), sought to avoid such an

Given the above approach, the empirical identification of over-education in individual-level survey data requires two measures: the level of education (or highest qualification) completed by each individual and the required education level of each individual's current job. Information on completed education is common in such datasets, though its accuracy is an important concern because mismeasurement may lead to error in the identification of over-education.[34, 35] Information on the required education levels of jobs, however, is not so common. In response, the over-education literature has developed several methods for estimating such a measure. The following sub-section discusses these methods, and establishes the preferred method.[36]

2.3.1 Methods for estimating required education levels of jobs

In the over-education literature, the required education levels of jobs have been estimated using three main methods: the *worker self-assessed (WA)*, *job analyst (JA)* and *realised matches (RM)* methods. Details of each method are discussed below.

Worker self-assessed (WA) method

The WA method uses individuals' responses to survey questions regarding the level of education required to perform their job. In some studies, the survey questions used have referred to the education required to *get* rather than *perform* the job. This subtle difference can lead to inconsistencies in the estimates of over-education across studies. In particular, if firms specify a higher level of education to obtain a job than is actually required to perform it, a phenomenon referred to as *credentialism* (Green et al., 1999; Borghans and De Grip, 2000), then studies based on survey questions that use the term *get* will tend to under-estimate the incidence of over-education (as the reported required education levels are artificially inflated). Of course, such inconsistencies will only arise if individuals perceive a difference between survey questions based on the use of *get* and *perform*, and then alter their responses accordingly. Nevertheless, given the concept that is being estimated, the WA method should be based on survey questions that refer to the education level required to *perform* the job.

---

assumption and identified *genuinely* and *apparently* over-educated graduates, where the distinction was based on individuals' satisfaction with the match between their education and job (i.e., those in non-graduate jobs who were unsatisfied with the match were deemed *genuinely over-educated*, while those satisfied with the match were *apparently over-educated*). In addition, Jensen (2003) and Guironnet and Peypoch (2007) developed methods that account for individuals' wages when identifying over-education, and Quinn and Rubb (2006) developed a method that adapts the RM method (discussed below) for use with panel data (i.e., to enable required education levels of jobs to vary over time when using the RM method). Of these four alternative methods, however, only the Chevalier (2003) approach has been used in subsequent studies (see, for example, Chevalier and Lindley (2006), Verhaest and Omey (2006b) and Green and Zhu (2010)).

[34] This point is frequently overlooked in the over-education literature (Van der Meer, 2006; Verhaest and Omey, 2006b; Leuven and Oosterbeek, 2011).

[35] It is for this reason that Section 3.4.1 scrutinises the quality of the education data used in this study.

[36] Based on this approach, resultant estimates of over-education must be validated. Recall, such validation—empirical tests of whether individuals identified as over-educated have under-utilised human capital—is the focus of Chapter 5.

The WA method can also be based on survey questions that ask individuals whether they believe themselves to be over-educated (or under-educated) for their job or whether, compared to their completed education, they believe a lower (or higher) education level is required for the job. This is a somewhat different approach to that described above because individuals' responses to such questions directly identify instances of over-education, rather than identifying the required education levels of jobs, and, therefore, it negates the need for comparisons to individuals' highest education levels.

Job analyst (JA) method

The JA method uses required skill levels defined in occupation classifications. For such occupation classifications, professional job analysts systematically evaluate the jobs in a particular labour market in order to determine the skill level required to perform each job and, based on the similarity of their tasks and skill level required to perform them, to group jobs into occupations (i.e., sets of jobs with similar tasks and required skill levels) and then organise occupations into increasingly larger groups. The result is a structure with numerous, hierarchical levels. At each level, a required skill level is defined for each of the occupation categories. Typically, the required skill levels are measured in terms of levels of education and these, therefore, are the estimates of the required education levels of jobs derived using the JA method. Examples of occupation classifications are the Dictionary of Occupational Titles (DOT) developed by the US Employment Service for the US labour market and the Australian Standard Classification of Occupations (ASCO) developed by the ABS for the Australian labour market (Borghans and De Grip, 2000; Kler, 2005).

Realised matches (RM) method

The RM method is based on the actual matches between individuals and jobs in labour markets. Similar to the JA method, it uses an occupation classification to perform analyses in terms of occupations rather than jobs, but, instead of using the required skill levels defined in such occupation classifications, the RM method uses the observed distributions of education levels among individuals employed in the same occupation to estimate the required education levels. Typically, the required education level is defined as the range of education levels within one standard deviation of the mean of the distribution observed for each occupation. Hence, individuals with an education level one standard deviation or more above the mean education level for their occupation are deemed to be over-educated.[37] As an alternative approach, some studies have also used the modal education level in these distributions as the required education level (i.e., individuals with an education level above the mode for their occupation are over-educated) (see, for example, Kiker, Santos and Mendes de

[37] Meanwhile, individuals with education level one standard deviation or more below the mean are deemed under-educated and individuals with education level within the one standard deviation range of the mean are deemed well-matched.

Oliveira (1997), Bauer (2002) and Tsai (2010)). The reason for using the mode is discussed later. Given its derivation of estimates of required education levels, the RM method is also often referred to as the 'empirical method' or 'statistical approach' (Sloane, 2003; Verhaest and Omey, 2006b).

Choosing between WA, JA and RM methods: The preferred method

Among the studies to have used more than one method, it is commonly found that the three methods result in significantly different estimates of the incidence of over-education, and that there is little correlation between such estimates (Battu et al., 2000; Groot and Maassen van den Brink, 2000b; Sloane, 2003; Kler, 2005; McGuinness, 2006; Verhaest and Omey, 2006a; Verhaest and Omey, 2006b; Verhaest and Omey, 2006c). Each method, therefore, tends to identify different individuals as being over-educated. The results of Verhaest and Omey (2006b) most clearly demonstrated this. Based on the variations in the three methods discussed above, they developed six different estimates of required education levels and found that while 66 per cent of individuals were identified as over-educated by at least one measure, only 3 per cent were identified as over-educated by all six measures.[38, 39] This disagreement between methods is an important limitation of the over-education literature. It may mean over-education is merely an artificial phenomenon identified in data as a result of the assumptions made regarding labour markets and individuals' accumulation of human capital from education (i.e., assuming jobs have productivity ceilings and that all individuals acquire the same human capital from a given qualification). At the very least, it means only one of the three methods accurately estimates the incidence of over-education.[40] The choice of method—and determination of the preferred method—is, therefore, critically important.[41]

For any given study of over-education, the choice of method is, of course, dependent on the information available in the dataset being examined. Recall, the WA method requires individuals' responses to a survey question on either the education required to perform (or get) their job or whether they believe themselves to over-educated. The JA method requires the assignment of each

[38] The six methods used in Verhaest and Omey (2006b) were: (i) WA method using survey question on education required to *perform* job; (ii) WA method using survey question on education required to *get* job; (iii) WA method using survey question on whether individuals have more education than required for job; (iv) JA method; (v) RM method using range within one standard deviation of the mean; and, (vi) RM method using the mode education level.
[39] A similar result was found for the identification of under-education: 25 per cent were identified as under-educated by at least one measure, but less than 1 per cent were identified as under-educated by all six (Verhaest and Omey, 2006b).
[40] And, given previous studies are based on methods producing inaccurate estimates of over-education, it also means that some of the empirical evidence regarding over-education (e.g., over-education wage penalty estimates) may be invalid.
[41] Some studies, such as Rumberger (1987), Kiker et al. (1997) and Battu et al. (2000), have argued that since estimates of the over-education wage penalty are unaffected by the method used, then the lack of correlation between methods is not a serious concern (Hartog, 2000; McGuinness, 2006). Such an argument is, however, flawed (Leuven and Oosterbeek, 2011). For the study of over-education, accurately estimating its incidence is the primary objective; any measurement error in its identification will likely bias the results of further analyses. The argument is further weakened by the fact there is indeed evidence of the wage penalty varying by method (see, for example, Groot and Maassen van den Brink, 2000a; Rubb, 2003b; Kler, 2005; Verhaest and Omey, 2006c). Verhaest and Omey (2006c) also found the effects of over-education on job satisfaction, job mobility and participation in training significantly differed by method used.

employed individual's job to an occupation classification that defines a required education (or skill) level for each occupation category. The RM method also requires the assignment of jobs to an occupation classification, though it is not necessary for the classification to define required education (or skill) levels for occupations. Since most individual-level survey datasets contain information based on an occupation classification, then, in terms of data requirements, the RM method is the least onerous of the three methods. For this reason, it is the method most often used in the literature.

The RM method, however, has many weaknesses. Its principal weakness is that it does not establish required education levels based on the requirements of jobs (i.e., their tasks and the skill level required to perform them). Instead, it uses the individual-job matches realised in labour markets and labels individuals over-educated if their education level is (one standard deviation or more) above the average in their occupation. Given its analysis of individual-job matches, the RM method uses information resulting from the interactions between labour supply and demand to estimate required education levels—a factor determined exclusively on the demand side. As a result, it is unlikely to produce reliable estimates of required education levels (Hartog, 2000; Verhaest and Omey, 2006a). Ultimately, the approach is inconsistent with the definition of over-education—individuals having human capital from education in excess of that required to perform their current job—and it is distinct from the other two methods. Hence, as argued by Hartog (2000), estimates of over-education based on the RM method should be regarded as capturing something quite different to those of the WA and JA methods and, as a result, they should not be compared.

Further weaknesses arise from the practical implementation of the RM method. The first concerns its use of an occupation classification and the need to assume that all jobs in the same occupation category have the same required education level. Such an assumption is, of course, more plausible at more detailed (or disaggregated) levels of the classification. But, in practice, sufficiently large numbers of individuals are needed in each occupation category to establish the distributions of education levels. This may prevent the use of a detailed level of the occupation classification and, therefore, adversely affect the plausibility of the homogeneity of jobs assumption (Sloane, 2003; McGuinness, 2006). Another weakness is its use of the range of education levels within one standard deviation of the mean as the required education levels. Since one standard deviation cut-off points are entirely arbitrary, the resultant estimates of the incidence of over-education are also arbitrary and, in fact, they are a negative function of the level of education within each occupation category. For instance, if an occupation contains a high proportion of truly over-educated individuals, then this would lead to a high mean education level and cut-off point for the occupation and, as a result, the RM method would under-estimate the true incidence of over-education within the occupation category (Borghans and De Grip, 2000; McGuinness, 2006). The one standard deviation cut-off

points also tend to impose symmetry on resultant estimates of over-education and under-education.[42] This symmetry is considered unrealistic: there is no reason to expect equality in the numbers of individuals over-educated and under-educated, and estimates based on the WA and JA methods do not exhibit such symmetry (Hartog, 2000; Sloane, 2003; McGuinness, 2006). It was the desire to avoid such symmetry that led to the modal education level within occupations being used as estimates of the required education levels.

The relative nature of the RM method—its use of the means and standard deviations of observed distributions—also means that it will *always* identify instances of over-education, regardless of whether it actually exists (Green et al., 1999; Jensen, 2003). This is a critical weakness. It also means the RM method is not suitable for analyses concerning the dynamics of over-education (Hartog, 2000; McGuinness, 2006; Leuven and Oosterbeek, 2011). In particular, its estimates are not suitable for examining how the incidence of over-education changes over time, as the RM method will tend to identify the same proportion of individuals as over-educated in each year. Furthermore, such estimates are not suitable for examining individuals' changes in over-education status over time (i.e., entries and exits from over-education) because it is possible for an individual's over-education status to change without any actual changes to their job or education level. For example, based on a decrease in the mean education level within their occupation, an individual can be identified as moving from well-matched to over-educated despite their job and level of education remaining unchanged. This potential for artificial variation also means the RM method is not suitable for analyses seeking to use the longitudinal variation in individuals' over-education status to identify (causal) parameters in an econometric model (e.g., using a fixed effects estimator to control for the effects of unobserved individual heterogeneity in an earnings function).

The WA method has three main strengths. The first is that survey questions refer specifically to each individual's particular job, and so there is no need for an aggregation of jobs into occupations and, as a result, no need for the homogeneity of jobs assumption (required in RM and JA methods). Individuals' responses also reflect the current requirements of their job. This means the WA method can capture changes over time in the required education levels of jobs and, therefore, its estimates can be used to examine the dynamics of over-education (Hartog, 2000). Additionally, compared to the costs associated with creating occupation classifications (as required in the RM and JA methods), it is a relatively cheap way in which to identify over-education (Hartog, 2000). There is, however, one key weakness in using the WA method. It is that individuals' responses to survey questions are entirely subjective: the assessment by one individual of the required education level of their job may

[42] For example, if the education levels within occupation categories follow roughly Normal distributions, then 68 per cent of the population lies within the range one standard deviation around the mean and, therefore, the RM method would tend to identify roughly 68 per cent of individuals as well-matched, 16 per cent over-educated and 16 per cent under-educated.

not be the same as that of another individual employed in the same job (Green et al., 1999).[43] Thus, the WA method captures whether individuals perceive themselves as over-educated for their job, but does not necessarily capture whether individuals—based on an objective evaluation of their human capital and that required for their job—are actually over-educated. The reliance on individuals' responses can also lead to biased estimates of over-education because individuals may: exaggerate the requirements of their job (to increase its status); merely report the hiring practices of their employer (a potential problem if credentialism exists or if employers adjust hiring practices in response to increases in the level of education within the labour force without adjusting job requirements); or, be less likely to respond to survey questions on employment, or surveys in general, if over-educated (Green et al., 1999; Borghans and De Grip, 2000; Hartog, 2000; Sloane, 2003; McGuinness, 2006). Such factors would upwardly bias the required education levels of jobs and, therefore, lead the WA method to under-estimate the incidence of over-education.[44]

The JA method, in contrast to the other methods, produces estimates of required education levels that are explicitly based on the requirements of jobs and objectively defined (i.e., derived independent of individuals and firms) (Rumberger, 1987; Hartog, 2000; Verhaest and Omey, 2006a). As a result, the estimates are consistent with the concept needed to identify over-education, consistent across individuals employed in the same job and unlikely to be biased. In these respects, the JA method is clearly superior to the WA and RM methods.[45] It is, however, not without its weaknesses. The main weakness is that occupation classifications are created in reference to a specific point in time—the time at which job analysts evaluated jobs—and so, unless updated, they cannot capture changes in job requirements that occur over time (Rumberger, 1987; McGuinness, 2006).[46] Since information on job requirements will eventually become out-dated, particularly for jobs that change rapidly due to technological advances, the length of time between the creation of the occupation classification and the period being examined is therefore critical to the validity of resultant over-education estimates (Borghans and De Grip, 2000; Chevalier, 2003). Old occupation classifications are likely to under-estimate required education levels and, as a result, over-estimate the

[43] A related issue here is that (some) individuals may be unable to accurately determine the level of education that provides individuals with the human capital necessary to perform the job.
[44] As previously discussed, the use of survey questions that refer to the education level required to *get*, rather than *perform*, the job can also lead to (downwardly) biased estimates of the incidence of over-education.
[45] The JA method may also account for the potential substitution of work experience (and on-the-job training) for formal education (Hartog, 2000). In particular, occupation classifications may quantify the required skill levels of jobs in terms of education levels and the years of (relevant) work experience that can substitute for such education. Recall, this affects the identification of under-education, but not over-education. Specifically, accounting for such a substitution is an attempt to minimise the number of (seemingly anomalous) cases where individuals are identified as having insufficient human capital to perform their job. Again, to the author's knowledge, no previous studies have accounted for this potential substitution.
[46] Hartog (2000), however, warned that updates of occupation classifications may merely reproduce earlier analyses and contain very little new information—specific reference was made to the US DOT and its updates. Hence, it cannot be assumed that updates account for changes in job requirements and contain revised estimates of required skill levels.

incidence of over-education.[47] A further implication, when using estimates based on the JA method to examine the dynamics of over-education, is that it must be assumed the required education levels of jobs remain fixed throughout the time period examined.[48]

It has also been argued that job analysts, based on their visits to employment sites, may not be able to ascertain all the information necessary to evaluate jobs, and that there may be some subjectivity in the assessments as job analysts may not always agree (Verdugo and Verdugo, 1992; Hartog, 2000; Van der Meer, 2006; Verhaest and Omey, 2006a). For instance, if some job analysts tend to define required skill levels based on education levels they observe among the individuals in each job or the hiring standards of the employers, rather than the requirements of jobs, then this will be a source of bias in the JA method (Hartog, 2000). Such factors would adversely affect the accuracy of the required skill levels defined in occupation classifications. Another potential weakness is that some occupation classifications do not measure required skill levels in terms of levels of education (Halaby, 1994; Hartog, 2000). Instead, mutually exclusive skill levels are defined (and simply labelled using an ordinal number scale, for example, 1 to 5) without any reference to the education levels necessary for individuals to perform jobs at each skill level.[49] The estimation of required education levels for each skill level requires specific assumptions, and as these vary across studies using the same occupation classification then so too will estimates of over-education.

In applying the JA method it is also necessary to assume the homogeneity of jobs within the same occupation category (Rumberger, 1987; Halaby, 1994; Green et al., 1999; McGuinness, 2006). In particular, it must be assumed that jobs assigned to the same occupation category have similar sets of tasks and, as a result, the same required education level. Recall, this assumption is more plausible at more detailed (or disaggregated) levels of the occupation classification. (It is also a requirement of the RM method, although an important difference when using the JA method is that there is no competing need for a sufficiently large number of individuals in each occupation category.) Thus, the most detailed level of the occupation classification (available in the dataset examined) can be used to maximise the likelihood that this homogeneity of jobs assumption holds. A further concern in the application of the JA method is the accuracy with which the dataset examined has coded individuals'

[47] Evidence in Van der Meer (2006) supports such a claim. In particular, based on a study of the Dutch labour market, Van der Meer (2006) used an occupation classification developed in the 1970s to estimate required education levels of occupations in the 1990s and found that resultant estimates of over-education were significantly smaller, and ultimately invalid, compared to estimates based on an occupation classification developed in the 1990s.

[48] Such an assumption is problematic if a broad time period is examined (e.g., a ten-year period) or if, over the period examined, it is likely that technological advances have heterogeneous effects on job requirements (i.e., if the degree to which required education levels change varies across jobs).

[49] A prominent example is the DOT: it defines required skill levels using the General Educational Development (GED) scale—an ordinal scale from 1 to 7—but, in empirical studies, there is no consensus on how to convert the GED scale to education levels (or years completed) (Halaby, 1994; Hartog, 2000; Leuven and Oosterbeek, 2011).

jobs to the occupation classification.[50] That is, to accurately identify over-education it is not only necessary for the required education levels in the occupation classification to be correctly defined, but, in the given data, the individuals must also be coded to the appropriate occupation category of the classification.[51]

Based on the above discussion of the three methods for estimating required education levels, it is clear the RM method, despite its widespread usage, is the least desirable method, and its estimates may not be suitable for comparison to those of the other methods (Hartog, 2000; Chevalier, 2003). Since the WA and JA methods both produce estimates consistent with the desired concept—required education levels that are based on job requirements—they both appear to be viable options. In choosing between the two, however, it is the objective nature and lower potential for biased estimates that lead to a preference for the JA method (Rumberger, 1987; Borghans and De Grip, 2000; Hartog, 2000; Verhaest and Omey, 2006a).[52] Of course, this preference is reliant on the existence and accessibility of a suitable occupation classification, whereby its suitability depends on: the degree of precision in its development; the amount of time between its creation and the period examined; the level of detail (or disaggregation) at which it can be used (in a given dataset); and, the accuracy with which individuals' jobs are coded to it (in a given dataset) (Hartog, 2000).

2.4 Over-education wage penalty estimates

Almost all studies in the over-education literature have considered the relationship between over-education and individuals' wages. In examining this relationship, most studies followed the approach of Duncan and Hoffman (1981): using cross-sectional data for individuals, a measure for years of required education is introduced to a Mincer (1974) earnings function to divide completed education into years in excess of that required (S_i^o), years required (S_i^r) and years less than that required (S_i^u). The resultant earnings function, which the literature has referred to as the ORU (Over-education, Required education and Under-education) earnings function, is:

$$\log w_i = X_i'\beta + \delta_o S_i^o + \delta_r S_i^r + \delta_u S_i^u + \varphi_1 Exp_i + \varphi_2 Exp_i^2 + \varepsilon_i \qquad [2.1]$$

where w_i is current (hourly) wage, X_i is a vector of controls for individual characteristics, Exp_i is a measure of labour market experience and ε_i are the usual regression disturbances. An alternative specification, which was developed in subsequent research, replaces the years of over-education and

[50] Although not mentioned earlier, this is also a concern for the application of the RM method.
[51] It is for this reason that Section 3.4.2 scrutinises the quality of the occupation data used in this study.
[52] Evidence from Verhaest and Omey (2006a), which performed 'encompassing tests' on estimates from the three methods, supports this conclusion. Specifically, it was found the WA and JA methods should always be preferred to the RM method, while results were ambiguous for preference between the WA and JA methods. However, based on the higher chance of bias in using the WA method, Verhaest and Omey (2006a) concluded the JA method is ultimately the preferred method.

under-education measures with indicators for being over-educated (O_i) and under-educated (U_i) and controls for completed education (S_i):

$$\log w_i = X_i'\beta + \lambda_o O_i + \lambda_u U_i + \gamma S_i + \varphi_1 Exp_i + \varphi_2 Exp_i^2 + \varepsilon_i \quad [2.2]$$

Such earnings functions are then estimated using the ordinary least squares (OLS) estimator.

Among the large number of studies that have estimated these ORU earnings functions, there are several consistent findings. In particular, based on [2.1], returns to years of over-education (δ_o) are found to be positive and statistically significant, though smaller in magnitude than returns to years of required education (δ_r), and then, based on [2.2], the effect of being over-educated (λ_o) is found to be negative and statistically significant (see Hartog (2000) and McGuinness (2006) for surveys of these results and Groot and Maassen van den Brink (2000a), Rubb (2003b) and Leuven and Oosterbeek (2011) for meta-analyses of the estimates).[53] These findings appear robust as they have emerged from studies examining various data sources, countries with varying education systems and labour market institutions, economies at different stages of development and using each of the three methods for identifying over-education (Rumberger, 1987; Kiker et al., 1997; Battu et al., 2000; Mendes de Oliveira, Santos and Kiker, 2000; Hartog, 2000; McGuinness, 2006).[54] Overall, these results indicate over-educated individuals earn more than their co-workers who have the (lower) required education for the job, but less than individuals with the same education who are employed in well-matched jobs (McGuinness, 2006; Sloane, 2007). It is this latter effect—being over-educated *causes* an individual's wage to be lower than if they were instead well-matched—that is referred to as the *over-education wage penalty*. Given OLS estimation of the above ORU earnings functions, the over-education wage penalty has been found to be, on average, roughly 15 per cent (though estimates have ranged between 8 and 27 per cent) (McGuinness, 2006).

Such OLS estimates, however, are likely biased. For these estimates to represent causal effects, it must be assumed that no factors correlated with both an individual's likelihood of being over-educated and their wage are absent from the model—this is the *conditional independence assumption*. Of most concern here are factors associated with individuals' stock of human capital—their skills from education, work experience, on-the-job training, innate abilities and physical and mental health—as each factor is likely correlated with both over-education and wages. While data used in previous studies contained measures for some of these factors, none had sufficient information to

---

[53] Estimation of [2.1] has also yielded the following results: returns to years of required education (δ_r) are positive, statistically significant and larger in magnitude than returns to completed education (based on comparison to the parameter for completed education in a standard Mincer earnings function); and, returns to years of under-education (δ_u) are negative, (usually) statistically significant and smaller in magnitude than δ_o and δ_r. Also, based on estimation of [2.2], the effect of being under-educated (λ_u) is found to be positive and statistically significant.

[54] A few studies have found the results can vary by method used to identify over-education, but it is typically the RM method that has produced significantly different results and this may be expected given its key differences compared to the JA and WA methods (Groot and Maassen van den Brink, 2000b; Rubb, 2003b; Kler, 2005; Verhaest and Omey, 2006c).

control for all of them. The main difficulty, of course, is controlling for individuals' innate abilities.[55] For this reason, several studies have considered whether innate ability is indeed correlated with the likelihood of being over-educated. The resultant evidence has indicated over-education is highly concentrated among individuals in lower and middle sections of the ability distribution and, therefore, innate ability appears negatively correlated with over-education (Green et al., 1999; Chevalier, 2003; McGuinness and Bennett, 2007; Korpi and Tåhlin, 2009).[56] Since innate ability is almost certainly also correlated with wages, then OLS estimates of δ_o and λ_o are biased due to the endogeneity of the over-education measures (i.e., unobserved individual heterogeneity results in non-random selection into over-education). Hence, these estimates cannot be interpreted as the causal effect that over-education has on individuals' wages.

Some studies have recognised this endogeneity problem. In attempts to overcome it, they have typically used longitudinal data to estimate ORU earnings functions that explicitly account for unobserved individual heterogeneity, such as:

$$\log w_{it} = X_{it}^{'}\beta + \lambda_o O_{it} + \lambda_u U_{it} + \gamma S_{it} + \varphi_1 Exp_{it} + \varphi_2 Exp_{it}^{\ 2} + \alpha_i + \varepsilon_{it} \qquad [2.3]$$

where α_i is a time-invariant random variable representing unobserved effects for individual *i*. The fixed effects estimator is then used to eliminate α_i and, thereby, produce unbiased estimates of the causal effect of over-education.[57] Evidence from this small body of research is mixed: some studies found the over-education wage penalty remained significant and in the range of 5 to 20 per cent (Dolton and Silles, 2008; Lindley and McIntosh, 2008; Korpi and Tåhlin, 2009; Verhaest and Omey, 2009; Mavromaras, McGuinness, Richardson, Sloane and Wei, 2011), while others found it declined to insignificance (i.e., a difference either not statistically significant or less than 1 per cent in size) (Bauer, 2002; Frenette, 2004; Tsai, 2010).[58] Regardless, each found fixed effects estimates of the wage penalty were significantly smaller than (pooled) OLS estimates, which was interpreted as empirical evidence confirming the endogeneity bias in OLS estimates.[59]

---

[55] This is similar to the *ability bias problem* encountered in the returns to education literature (see, for example, Card (1999), Blundell, Dearden and Sianesi (2004) and Angrist and Pischke (2009)); it is also often referred to as the *selection problem* or *omitted variables problem*.

[56] One study with evidence to the contrary, however, is De Grip, Bosma, Willems and van Boxtel (2008): it found the over-educated to be no less able, at least in terms of cognitive abilities, than well-matched individuals with the same education.

[57] Some studies also used the random effects estimator, but, since the necessary assumption that α_i be uncorrelated with X_{it} is likely violated, these estimates are likely biased. Hence, they are not considered here. This no correlation assumption is further discussed in Section 6.4.1. Also, the fixed effects estimator is further discussed in Section 5.3.2.

[58] As in this study, Mavromaras et al. (2011) used the HILDA Survey data for the 2001–2008 period to derive wage penalty estimates. Results similar to Mavromaras et al. (2011) are also reported in Mavromaras, McGuinness, O'Leary, Sloane and Wei (2010c) and Jones, Mavromaras, Sloane and Wei (2011).

[59] Dolton and Silles (2008) was an exception: it used instrumental variables methods to also account for the potential measurement error in the required education measure and, ultimately, concluded that upward bias from unobserved individual heterogeneity was offset by an equal downward bias from measurement error. Hence, it argued OLS estimates actually represent causal effects.

For each of these studies, however, data limit the validity of the estimates. In particular, since fixed effects estimates are identified by the individuals in the data who move between over-educated and well-matched states, then the variation in over-education status is critical. This has two main implications. First, this variation must be valid: it must represent individuals who move from being over-educated to well-matched, or vice versa. The method used to identify over-education is therefore a relevant concern. Recall from Section 2.3, the relative nature of the RM method means it is not a suitable for considering changes in over-education status and, as a result, it should not be used in deriving fixed effects estimates of the over-education wage penalty. The validity of estimates from Bauer (2002), Lindley and McIntosh (2008), Tsai (2010) and Mavromaras et al. (2011) are consequently limited given each study used the RM method.[60] Second, for estimates to represent causal effects, the individuals whose over-education status changes must be representative of all over-educated individuals (i.e., there must be no selection effects in the variation in over-education status). As a result, data where few individuals are observed changing over-education status are unlikely to produce valid causal effects estimates and, therefore, the year-to-year variation in individuals' over-education status and the lengths of the panels (i.e., number of observations per individual) in the data analysed acutely affect the validity of the fixed effects estimates.[61] In these respects, almost all of the above studies are deficient: Frenette (2004), Dolton and Silles (2008), Korpi and Tåhlin (2009) and Verahest and Omey (2009) are based on data with short panels, while the data used in Bauer (2002) and Tsai (2010) contained little variation in over-education status.[62, 63] Given the data limitations, it is difficult to interpret the above fixed effects estimates as evidence on the causal effects of over-education.

A few studies have also attempted to overcome the endogeneity problem by using semi-parametric matching methods to estimate the over-education wage penalty; specifically, propensity score matching estimators were used. These estimates are essentially derived by comparing the wages of over-educated individuals with those of well-matched individuals who are otherwise identical. However, similar to regression analyses using cross-sectional data, this approach relies on the conditional independence assumption to identify causal effects, and so it is not guaranteed to

[60] With regards to estimating the over-education wage penalty, a further limitation of Mavromaras et al. (2011) is that it simultaneously considered the over-skilling issue and categorised individuals as: 'over-educated only'; 'over-skilled only'; and, 'over-skilled and over-educated'. Thus, the estimated effects are not equivalent to the over-education wage penalty as defined in this study. Recall, the over-skilling issue was (briefly) discussed (in a footnote) in Section 2.3.

[61] Generally, in applied economics research, a minimum of 5 observations per individual is considered necessary to produce valid fixed effects estimates (Baltagi, 2001; Wooldridge, 2003; Cameron and Trivedi, 2005). The number of individuals whose over-education status changes will, of course, also affect the precision of the fixed effects estimates.

[62] Specifically, Frenette (2004) and Dolton and Silles (2008) used data with two observations per individual, Verhaest and Omey (2009) three observations and Korpi and Tåhlin (2009) four observations. Meanwhile, year-to-year variation in over-education status is between 5 and 16 per cent in Bauer (2002), which the author acknowledged cast doubt on the fixed effects estimates, and only around 3 per cent in Tsai (2010).

[63] Lindley and McIntosh (2008) and Mavromaras et al. (2011) did not report the year-to-year variation in over-education.

overcome the endogeneity problem; a point overlooked in these studies. It is, therefore, not surprising these studies found wage penalty estimates either similar to OLS estimates (McGuinness, 2008; McGuinness and Sloane, 2011) or only slightly reduced (Lamo and Messina, 2010).[64]

Overall, as highlighted by Leuven and Oosterbeek (2011), because the over-education literature has largely overlooked identification issues that have been prominent in recent labour economics and economics of education research, most of the wage penalty estimates cannot be interpreted as causal effects.[65] Further, among the few studies to have considered these identification issues and attempted to overcome the endogeneity problem, data limitations are such that the resultant estimates of causal effects are unlikely to be valid. To date, therefore, there is no clear evidence of an over-education wage penalty.

2.5 Dynamics of over-education

As mentioned in Section 2.2, there is little empirical evidence on the dynamics of over-education. That is, few—if any—studies have examined issues such as: individuals' entries to and exits from the over-educated state; the persistence (or duration) of over-education for individuals; how its effects evolve for individuals who remain over-educated (e.g., its effects on wages and job satisfaction levels by individuals' durations over-educated); and, whether there are enduring effects for individuals who are no longer over-educated (e.g., prior over-education affecting the wages of currently well-matched individuals). Evidence on the dynamics of over-education is critical for determining whether it is a serious problem in modern labour markets—recall, if it is short-term disequilibria with no enduring effects, then over-education could be regarded as just a by-product of dynamic and well-functioning labour markets—and, as a result, consideration of such dynamics has been identified as a key avenue for future research (see, for example, Dolton and Silles (2003), Pollmann-Schult and Büchel (2004), Green and McIntosh (2007), McMillen, Seaman and Singell (2007), European Centre for the Development of Vocational Training (CEDEFOP) (2009) and Leuven and Oosterbeek (2011)).

The following sub-sections discuss existing evidence on the dynamics of over-education. In particular, Section 2.5.1 discusses evidence on the persistence of over-education for individuals,

[64] Two other approaches have been used to try and overcome the endogeneity problem, though both involved the estimation of ORU earnings functions. First, studies have restricted their analyses to more homogenous samples (e.g., recent cohorts of graduates) to minimise the potential for unobserved differences between individuals; they found the over-education wage penalty remained significant (Battu et al., 2000; Dolton and Vignoles, 2000; Kler, 2005; McGuinness and Bennett, 2007). The second approach used estimated residuals from a model for initial (post-graduation) wages to proxy unobserved individual heterogeneity in a model for current wages; they also found the wage penalty remained significant (Chevalier, 2003; Chevalier and Lindley, 2006).

[65] Leuven and Oosterbeek (2011) argued the endogeneity bias may be of such size and significance that it entirely explains the wage differences observed between over-educated and well-matched individuals (i.e., no over-education wage penalty exists). Also, in addition to the endogeneity bias, it was argued that estimates may contain attenuation bias due to measurement error in completed education and required education measures. In response, this study exerts considerable effort in ensuring the accuracy of such measures. This is further discussed in Section 3.4.

while Section 2.5.2 considers results from studies that have examined whether over-education has enduring effects for individuals—specifically, such research has focused on the possibility over-education leads to human capital depreciation among individuals.[66]

2.5.1 Persistence for individuals

Most empirical evidence on the persistence of over-education is based on simple cross-tabulations of individuals' over-education status between two points in time (i.e., the number over-educated at both points reported as a proportion of those initially over-educated). Such evidence, however, is mixed. Findings from some studies suggest over-education is highly persistent. For example: Battu, Belfield and Sloane (1999) found 85 per cent of UK graduates over-educated one year after graduation were also over-educated six years after graduation, and 75 per cent were also over-educated eleven years after graduation; Frenette (2004) found 74 per cent of Canadian graduates over-educated two years after graduation were also over-educated five years after graduation; and, De Grip et al. (2008) found 90 per cent of all the individuals over-educated in the Netherlands in the early 1990s were also over-educated six years later.[67] Meanwhile, results of other studies suggest much less persistence in over-education: Dorn and Sousa-Poza (2006) found 45 per cent of all the individuals over-educated in Switzerland in 1999 were also over-educated one year later, 20 per cent two years later, 11 per cent three years later and only 8 per cent four years later; and, for the UK, Lindley and McIntosh (2008) found 44 per cent of all the individuals over-educated in 1991 were also over-educated five years later, 32 per cent ten years later, 26 per cent fourteen years later and, overall, only 16 per cent were over-educated at each point in time.[68]

Several factors may explain this mixed evidence regarding persistence. First, differences may arise because different samples were analysed—Battu et al. (1999) and Frenette (2004) examined recent university graduates, while the other studies examined all individuals—and because the degree of persistence varies across the population. That is, the evidence of persistence in Battu et al. (1999) and Frenette (2004) may arise because over-education is more persistent among recent graduates than it is, on average, among all employed individuals. Moreover, the results of De Grip et al. (2008) may overstate the persistence of over-education because the sample was restricted to individuals employed at both points in time. This restriction means that, compared to the estimates from the other studies, the individuals who are not employed at the second point in time have been removed

[66] Evidence from studies that examined entries and exits from over-education—Groot and Maassen van den Brink (2003), Frenette (2004), Bender and Heywood (2006), Chevalier and Lindley (2006), Dorn and Sousa-Poza (2006), Robst (2007) and Lindley and McIntosh (2008)—is not discussed here because, due to the basic descriptive and multivariate analyses performed and significant data limitations, it provides few valuable insights (e.g., it merely found individuals with higher education more likely to enter over-education and older individuals less likely to exit). To the author's knowledge, no study has examined how the effects of over-education evolve over individuals' durations over-educated.

[67] Robst (1995), Chevalier (2003) and Rubb (2003a) also contain evidence of persistent over-education.

[68] Dolton and Silles (2003) and Groot and Maassen van den Brink (2003) also contain evidence of transient over-education.

from the denominator of the calculation, which thereby inflates the proportion reported as persistently over-educated. The differences may also arise because different methods were used to identify over-education. In particular, since Lindley and McIntosh (2008) used the RM method, which, recall from Section 2.3, is not suitable for examining individuals' changes in over-education status over time, it may be inappropriate to compare their results to those of the other studies. It is also possible that the differences in results are due to cross-country differences regarding over-education and its persistence.

Along with the resultant mixed evidence, a limitation of these simple analyses is that for the reported proportions to be evidence of persistence it must be assumed individuals are over-educated for the entire period of time between the two points examined. If not the case, and individuals instead spend time well-matched (or under-educated), then it is difficult to argue that such evidence is indicative of persistence as the total time spent over-educated is unknown. The assumption is, of course, more plausible for points closer together; studies using data points farther apart than two years arguably produce unreliable evidence.[69, 70] On this basis, Dorn and Sousa-Poza (2006) (and Groot and Maassen van den Brink (2003) and Rubb (2003a)) provide the only reliable evidence. This limitation reveals the overriding weakness of the existing research, which, as highlighted by Pollmann-Schult and Büchel (2004), is that it has not examined the factor most relevant to the persistence of over-education: the durations of individuals' spells over-educated. The weakness is due to data limitations. Specifically, information on individuals' durations over-educated was not available in the data analysed and, since most studies were based on either cross-sectional data with (some) recall information or longitudinal data with a short panel-element, such information was unable to be approximated. Verhaest and Schatteman (2010) is somewhat of an exception: based on monthly data for a seven-year period around the early-2000s in Flanders, it found roughly 30 per cent of recent school leavers (who did not proceed to tertiary education) were over-educated for almost the entire period, and a further 20 per cent combined over-education with long periods of joblessness. Such evidence is much closer to estimating individuals' durations over-educated, but is by no means a definitive representation of the persistence of over-education throughout labour markets (as it merely reflects experiences of a small group of labour market participants in a specific, seven-year period).

Overall, given the limitations associated with the empirical evidence, it is difficult to draw any firm conclusions regarding the persistence of over-education for individuals. At the very least,

[69] This assertion is based on empirical evidence regarding the frequency with which individuals change jobs (or occupations) as such transitions affect their likelihood of remaining over-educated between the two points; for example, as discussed in Section 3.4.2, approximately 20 per cent of individuals employed in Australia change occupation over one-year intervals.
[70] Elias and Purcell (2004) and Verhaest and Schatteman (2010) recognised this limitation and, in response, presented graphical plots of the incidence of over-education (in the UK and Flanders) on a monthly basis over seven-year periods. However, since the plots are not conditional on previous over-education status, they do not indicate whether it is the same individuals over-educated over time and, therefore, they provide no evidence regarding the persistence of over-education.

however, it would appear reasonable to conclude that over-education is not an entirely temporary phenomenon. In particular, it appears persistent for recent labour market entrants—recent university graduates and recent school leavers without tertiary qualifications—with some also over-educated around ten years after being initially observed over-educated.[71, 72]

2.5.2 Link to human capital depreciation

Human capital depreciation (or skill obsolescence) is generally considered to result from skill atrophy following extended periods where individuals' skills are insufficiently used; prime examples are periods of unemployment and non-participation in labour markets (e.g., career interruptions to raise children) (Mincer and Polachek, 1974; Mincer and Ofek, 1982; Pissarides, 1992).[73] Since individuals' human capital appears to be under-utilised while over-educated, it is possible periods of over-education may lead to human capital depreciation. Only two studies have examined such a possibility: Rubb (2006) and De Grip et al. (2008). But, this is not surprising as, in general, there has been limited research conducted into human capital depreciation.[74] Rubb (2006) argued that if over-education led to human capital depreciation, then the over-educated would earn lower returns to labour market experience than well-matched individuals. Thus, based on US Current Population Surveys (CPS) data for the 1994 to 2000 period, Rubb (2006) estimated ORU earnings functions, such as [2.1] in Section 2.4, with interactions between experience and the (years of) over-education measure. Coefficient estimates for these interaction terms were then used to empirically test for a link between over-education and human capital depreciation. Rubb (2006) found over-educated individuals were indeed earning statistically significantly lower returns to experience. Such estimates, however, are likely biased. This is because, as discussed in Section 2.4, ORU earnings functions estimated using cross-

---

[71] Studies evaluating predictions of career mobility theory, as discussed in Section 2.2, may also provide (indirect) empirical evidence on persistence. Recall, the theory predicts over-education occurs early in individuals' careers, and such individuals are more likely to receive on-the-job training, job promotions, greater wage growth and have higher rates of job mobility; hence, over-education is predicted to be only temporary for individuals. For each prediction, however, there is evidence to both support and contravene it. For instance, some studies found over-educated individuals more likely to be promoted or move to jobs requiring higher education levels (see, for example, Sicherman (1991), Hersch (1995), Dekker, De Grip and Heijke (2002), Groeneveld and Hartog (2004), Alba-Ramírez and Blázquez (2003) and McMillen et al. (2007)), while other studies found their job mobility did not lead to improved matches nor did they experience greater rates of wage growth (see, for example, Sloane et al. (1999), Büchel and Mertens (2004) and Korpi and Tåhlin (2009)). As a result, the evidence regarding career mobility theory provides little insight into the persistence of over-education.

[72] In the over-education literature, Dolton and Vignoles (2000) is frequently cited as providing evidence over-education is persistent for individuals. Such a conclusion, however, is unwarranted. This study examined UK graduates in the 1980s and found 38 per cent were over-educated in their first job and then six years later 30 per cent were over-educated. These are merely estimates of the incidence of over-education at two points in time and, therefore, provide no information on whether it is the same individuals over-educated over time. This point continues to go unrecognised in the literature.

[73] The ageing process (e.g., deterioration in physical or mental health) and labour market changes that affect the value of human capital (e.g., skill-biased technological change) may also result in human capital depreciation (De Grip, 2006). In this study, however, the concern is for depreciation resulting from skill atrophy.

[74] De Grip and Van Loo (2002) and De Grip (2006) provide reviews of the existing literature, and similarly conclude there is little empirical evidence regarding human capital depreciation. Mincer and Polachek (1974) and Mincer and Ofek (1982) represent early papers on the topic, while Albrecht, Edin, Sundström and Vroman (1999), Kunze (2002) and Görlich and De Grip (2009) are examples of more recent studies of human capital depreciation.

sectional data likely contain unobserved individual heterogeneity, particularly with respect to each individual's stock of human capital, which leads to non-random selection into over-education (i.e., endogeneity of the over-education measure) and, as a result, biased coefficient estimates.

De Grip et al. (2008) performed a more in-depth analysis of the potential link between over-education and human capital depreciation. In particular, they used data from the Maastricht Aging Study (MAAS) conducted in the Netherlands during the 1990s, where individuals completed cognitive abilities tests—a standard set of neuropsychological tests assessing their verbal memory, cognitive flexibility, verbal fluency and information processing speed—at two points in time, six years apart. Based on these test scores, models for the change in each individual's test scores were estimated with measures for being over-educated in the initial period included to empirically test whether over-education had led to any human capital depreciation. De Grip et al. (2008) found only weak evidence that it had: over-education led to a statistically significant decline in individuals' verbal fluency (though the effect was only statistically significant at the 10 per cent level), but had no effect for all other measures of cognitive abilities.[75] It is possible, however, De Grip et al. (2008) found limited evidence of human capital depreciation because they examined human capital in terms of individuals' general aptitudes—their cognitive abilities—which, when compared to job-specific human capital, may be less likely to atrophy while individuals are over-educated. For instance, for an over-educated individual who completed a Veterinary Science degree but works as a sales assistant, it is likely their human capital related to working as a veterinarian would atrophy much more quickly than their general aptitudes.

Overall, Rubb (2006) and De Grip et al. (2008) have provided no conclusive evidence that over-education leads to human capital depreciation among individuals. What is evident, however, is that this potential link warrants further consideration in the over-education literature.[76]

2.6 Individual preferences and over-education

As previously discussed, education has many potential benefits. Depending on individuals' preferences, these benefits can include various attributes of jobs, such as the wage, hours of work, job security, job flexibility and required effort. Despite its primary concern being the benefits of education, the over-education literature has, to date, focused on only one benefit: wages. Then, based

[75] Some further significant effects were found in analyses that used a proxy measure for the extent (or severity) of over-education (instead of the identifier for being over-educated): as expected, greater over-education (i.e., individuals with greater amounts of under-utilised human capital) led to greater declines in individuals' cognitive abilities.
[76] To the author's knowledge, no study has examined the possibility prior over-education affects the wages of currently well-matched individuals. Also, no study has tested for state dependence in over-education. (Mavromaras and McGuinness (2012) tested for such state dependence in over-skilling, but did not interpret it as evidence of human capital depreciation nor offer an explanation as to why it may exist.)

on estimated wage penalties, as discussed in Section 2.4, the literature has inferred all instances of over-education represent labour market failures. But, with a broader assessment of the benefits, this may not be true. Since many job attributes, not just the wage, can affect individuals' utility levels, some over-educated individuals may have actually obtained jobs that maximise their (expected) utility and, therefore, achieved their preferred outcome.[77] Recall, consistent with compensating wage differentials, these individuals trade wages for non-pecuniary benefits of employment or improved working conditions and, as a result, accept jobs for which they are over-educated. Hence, they are referred to as *voluntarily over-educated* in this study. By effectively treating individuals as wage maximisers, rather than utility maximisers, the literature has overlooked the role individuals' preferences may play in the incidence of over-education.

Few studies have acknowledged the possibility of voluntary over-education and, as a result, there is little empirical evidence on its existence. Nevertheless, relevant evidence can be found in studies that have considered over-education more generally. In particular, given the proposed method for identifying voluntary over-education—which, recall from Chapter 1, is a distinction based on over-educated individuals' job satisfaction levels and desire for a new job—then evidence on the effect being over-educated has on job satisfaction and intentions to quit is relevant. Based on the many studies that have examined the link between over-education and job satisfaction, the prevailing finding is that, compared to the well-matched, over-educated individuals are significantly less satisfied with their jobs (Tsang et al., 1991; Battu et al., 2000; Bender and Heywood, 2006; Verhaest and Omey, 2006b; Allen and de Weert, 2007; Amador, Nicolás and Vila, 2008; Fleming and Kler, 2008; Korpi and Tåhlin, 2009; Green and Zhu, 2010). In addition, several studies have found the over-educated are more likely to want to quit, be actively searching for a new job and actually experience voluntary job separations (Tsang et al., 1991; Hersch, 1995; Robst, 1995; Sloane et al., 1999; Groot and Maassen van den Brink, 2003; Wald, 2005; Di Pietro and Urwin, 2006; Allen and de Weert, 2007; McGuinness and Wooden, 2009). Such evidence suggests over-education is an involuntary state for most individuals, but does not preclude the possibility it is voluntary for some.[78]

Among the few studies to have considered voluntary over-education, there are two with relevant empirical evidence, both of which are very recent: McGuinness and Sloane (2011) and Mavromaras et al. (2011).[79, 80] Both studies supposed that any individuals whose job satisfaction levels

[77] This may also occur if individuals derive consumption (or lifestyle) benefits from being a student.
[78] Studies that model entry to over-education also have the potential to provide relevant evidence; Groot and Maassen van den Brink (2003), Frenette (2004) and Dorn and Sousa-Poza (2006) performed such analyses. However, given data limitations led to the estimation of relatively simple probability models, these studies provide little insight into voluntary over-education: they merely found individuals with higher education were more likely to enter over-education, while those who initially worked more hours and had more tenure were less likely.
[79] Results similar to Mavromaras et al. (2011) are also reported in Mavromaras et al. (2010c) and Jones et al. (2011).
[80] To the author's knowledge, only two further studies have examined the possibility of voluntary over-education: Chevalier and Lindley (2006) and Robst (2007). Both, however, are subject to two major limitations: first, they did not use standard

are not reduced by being over-educated could be regarded as voluntarily over-educated, with such instances likely the result of trade-offs between wages and other job attributes (i.e., a form of compensating wage differentials). Thus, each study began by examining the relationship between over-education and job satisfaction, and then examined the relationship between over-education and various job attributes (e.g., job security, work autonomy, hours and job flexibility). Both studies also simultaneously considered the issue of over-skilling (i.e., regression analyses contained controls for being over-skilled); in fact, Mavromaras et al. (2011) combined it with over-education and examined the following categories of *job mismatch* (with all compared to the well-matched): 'over-educated only'; 'over-skilled only'; and, 'over-skilled and over-educated'.

McGuinness and Sloane (2011), using the Flexible Professional in the Knowledge Society (REFLEX) data collected in 2005 from UK university graduates who had graduated five years earlier, found being over-educated significantly reduced individuals' job satisfaction levels. But since the reduction was smaller in magnitude than the effect of being over-skilled, they argued it was evidence of over-education containing a voluntary element. In further support of voluntary over-education, they also found a positive, statistically significant relationship between over-education and individuals' perceived job security.[81] Similarly, Mavromaras et al. (2011), based on analysis of the HILDA Survey data for the 2001–2008 period, found over-education resulted in reduced job satisfaction, with the effect mostly found among those who were also over-skilled (i.e., 'over-skilled and over-educated'). Thus, some of the 'over-educated only' may be voluntarily over-educated individuals. With respect to job attributes, Mavromaras et al. (2011) found evidence of over-education leading to increased satisfaction with job security for males with a university degree and increased satisfaction with job flexibility for females who only completed secondary education.[82] However, based on an overall assessment of the evidence—where the estimated wage and job satisfaction effects of being 'over-educated only', 'over-skilled only' and 'over-skilled and over-educated' were considered for males and females by highest education level, and it was argued that any groups experiencing wage penalties and increased (or unchanged) job satisfaction levels were

---

definitions of over-education (recall from Section 2.3, Chevalier and Lindley (2006) examined *apparent* and *genuine* over-education in attempts to identify over-educated individuals with under-utilised skills, whilst Robst (2007) examined mismatches in terms of field of study); and, second, both only examined university graduates. Such limitations make it difficult to interpret their results as direct evidence regarding voluntary over-education. In addition, their results are not particularly instructive; both merely conducted descriptive analyses of the reasons over-educated individuals accepted their jobs and found some evidence consistent with voluntary over-education: large proportions whose main reason was either better pay or promotion opportunities, career development, a change of interests or better working conditions.

[81] Ortiz (2010) similarly found over-educated individuals had greater (perceived) job security; this finding, however, was not interpreted as evidence of voluntary over-education.

[82] Specifically, Mavromaras et al. (2011) examined the effect over-education had on individuals' satisfaction with their pay, job security, hours worked, job flexibility (or work-life balance achieved) and the work itself. Fleming and Kler (2008) used the same data, though only for 2001, to perform similar analyses for males; they found the over-educated were significantly less satisfied with their jobs overall and less satisfied with their pay, job security, hours worked and the work itself. Since this study also uses HILDA Survey data, the same (proxy measures for) job attributes are examined in Chapter 7.

evidence of a compensating differential (or voluntary over-education)—Mavromaras et al. (2011) concluded there was only weak evidence of voluntary over-education. In fact, the only evidence consistent with voluntary over-education was found among females whose highest education was a Diploma or Advanced diploma and who were 'over-educated only': they were found to incur a wage penalty of roughly 5 per cent, but had no significant reduction in job satisfaction.

The results of McGuinness and Sloane (2011) and Mavromaras et al. (2011) are subject to some important limitations. First, the ordinal measures for job satisfaction and job attributes were dichotomised into binary variables and probit models estimated, rather than econometric models and estimators specifically designed for ordinal-valued measures.[83] This simplification, with its arbitrary choice of cut-points for dichotomisation and resultant loss of information, leads to inefficient and potentially biased estimates. Further, McGuinness and Sloane (2011) used cross-sectional data and univariate probit models that do not account for any effects of unobserved individual heterogeneity; these univariate probit models likely lead to upwardly biased estimates for the effect of over-education (Fleming and Kler, 2008). On the other hand, Mavromaras et al. (2011) accounted for such heterogeneity by using panel data to estimate random effects probit models (with Mundlak-Chamberlain controls). But a limitation of these results is that Mavromaras et al. (2011) used the RM method to identify over-education: recall, its relative nature makes it inappropriate for examining changes in over-education status, which, given the estimation of random effects models, is effectively what Mavromaras et al. (2011) has done.[84] The final, and most important, limitation of McGuinness and Sloane (2011) and Mavromaras et al. (2011) is that both did not distinguish between voluntarily and involuntarily over-educated individuals prior to considering job attributes for evidence of trade-offs. Instead, the effect over-education has on job attributes was estimated for all over-educated individuals combined (i.e., models were estimated using a single 'over-educated' identifier). Such analyses, however, are unlikely to produce evidence of voluntary over-education.[85] This is because voluntary over-education and involuntary over-education have opposite effects on job attributes—as the involuntarily over-educated are unlikely to experience compensatory improvements in job attributes—and the existing evidence suggests over-education is involuntary for most individuals. Thus, the estimated effects of over-education will tend to reflect the circumstances of involuntarily

[83] Specifically, McGuinness and Sloane (2011) used job satisfaction measures with a scale from 1 (very dissatisfied) to 5 (very satisfied) and deemed individuals 'satisfied' if they reported a value of 4 or 5, while job attribute measures had a scale from 1 (does not apply at all) to 5 (applies to a very high extent) and were dichotomised at a value of 5. Meanwhile, Mavromaras et al. (2011) used job satisfaction measures with a scale from 0 (totally dissatisfied) to 10 (totally satisfied) and deemed individuals 'satisfied' if they reported a value between 7 and 10.

[84] For the study of over-education, a further limitation of Mavromaras et al. (2011) is that it does not directly estimate the effect over-education has on job satisfaction and job attributes, rather it estimated the effects of so-called job mismatches.

[85] Recall, Mavromaras et al. (2011) estimated models with two identifiers related to over-education: 'over-educated only' and 'over-skilled and over-educated'. Nevertheless, the resultant estimates are subject to the limitation discussed here.

over-educated individuals (i.e., evidence of voluntary over-education will be overshadowed, or at least diminished, by the effects of involuntary over-education).

In the over-education literature, the prevailing view is that individuals are over-educated on an involuntary basis. This is a contention that has been rarely empirically tested. This shortcoming, along with the evidence, albeit limited, from the recent studies of McGuinness and Sloane (2011) and Mavromaras et al. (2011), suggests considering the potential for voluntary over-education is a suitable avenue for future research.

2.7 Over-education in Australia

As with other modern societies, there is empirical evidence to suggest over-education exists in Australian labour markets. This evidence is consistent with findings from the over-education literature more generally. In particular, estimates of the incidence of over-education, among the working-age employed individuals in Australia, have ranged between 15 and 30 per cent for males and between 10 and 25 per cent for females (Linsley, 2005; Voon and Miller, 2005; Kler, 2006; Messinis and Olekalns, 2007a; Fleming and Kler, 2008; Jones et al., 2011; Mavromaras et al., 2011).[86] Over-educated individuals earn more than their (less educated) co-workers, but less than individuals with the same education who are well-matched (Kler, 2005; Linsley, 2005; Voon and Miller, 2005; Kler, 2006; Messinis and Olekalns, 2007a; Mavromaras et al., 2010c; Jones et al., 2011; Mavromaras et al., 2011; Ryan and Sinning, 2011b). And, compared to the well-matched, over-educated individuals are less satisfied with their jobs (Kler, 2006; Fleming and Kler, 2008; Mavromaras et al., 2010c; Jones et al., 2011; Mavromaras et al., 2011). Similar to the evidence for other countries, however, this evidence is subject to key limitations. For instance, estimates of the incidence of over-education were found to vary depending on the method used to identify it (Kler, 2005), and most wage penalty estimates were derived using cross-sectional data and ORU earnings functions and are, therefore, likely biased. A further similarity is that little is known about the dynamics of over-education in Australia, particularly with respect to its persistence for individuals and its potential link to human capital depreciation.

The body of research considering over-education in Australia, when compared to the number of studies concerning other countries, is rather small, but it is an active field of research.[87] Evidence of this includes the policy forum on 'Education and Skill Mismatches in the Labour Market' that was published in *The Australian Economic Review* in 2007—consisting of Mavromaras and

[86] Previous estimates of over-education in Australia are further discussed in Chapter 4: they are presented in Table 4.4 and then compared to the estimates derived in this study.
[87] Roughly 20 studies have considered over-education in Australia; the broader over-education literature, however, consists of probably close to 200 studies, with most examining European countries and the US.

McGuinness (2007), Sloane (2007), Miller (2007), Messinis and Olekalns (2007b) and Mavromaras, McGuinness and Wooden (2007)—and which discussed much of the over-education literature and Australian evidence. There are also recent studies focusing on over-education among immigrants to Australia: Kler (2007); Green, Kler and Leeves (2007); Messinis (2008); Chiswick and Miller (2010); and, Piracha, Tani and Vadean (2010).[88] In addition, the National Centre for Vocational Education Research (NCVER) has published a series of research reports examining over-education and skills utilisation in Australian labour markets, with a particular focus on individuals who completed vocational qualifications.[89] These include Karmel (2009), Ryan and Sinning (2009, 2011a, 2011b), Mavromaras, McGuinness and Fok (2010a) and Mavromaras et al. (2011).[90, 91]

In some cases, studies examining Australian labour markets are also leading the advances in the broader over-education literature. Specifically, Mavromaras et al. (2010c), Jones et al. (2011) and Mavromaras et al. (2011) all sought to establish over-education wage penalty estimates that could be interpreted as causal effects. These studies, along with Fleming and Kler (2008), also examined the effect over-education had on individuals' job satisfaction levels, both overall and with respect to its various attributes, with the aim of determining whether some are voluntarily over-educated. Furthermore, Jones et al. (2011) is one of the first studies to have considered the potential link between disability and over-education. In each case, the innovation is, to a large extent, due to the availability of a quality longitudinal dataset for Australia: namely, the HILDA Survey data. This dataset has made it possible to examine individuals' behaviour over time and to consider previously unexplored factors regarding over-education.

Based on the existing evidence, over-education appears to be a significant economic issue in Australian labour markets. There are, however, many unanswered questions regarding over-education in Australia, and hence it remains an issue ripe for further research.

2.8 Conclusion

As discussed in this chapter, the over-education literature has several important limitations. There is the lack of consensus regarding a theoretical framework to explain the existence of over-education.

---

[88] These studies found immigrants, particularly those from a non-English speaking background (NESB), are more likely to be over-educated than native-born Australians. Over-education also appeared to have a significant impact on the wages of immigrants, with wage penalty estimates largest for NESB immigrants. The adverse experiences of immigrants, however, may not be entirely due to language difficulties, imperfect transferability of human capital and discrimination because Piracha et al. (2010) found prior over-education (i.e., being over-educated in home country prior to immigration) significantly affected the labour market outcomes of immigrants.

[89] In Australia, these are referred to as Vocational Education and Training (VET) qualifications.

[90] The key results from these studies have been discussed above and in previous sections.

[91] There are also recent studies considering the incidence and effects of over-skilling, rather than over-education, in Australia, such as Mavromaras et al. (2009), McGuinness and Wooden (2009), Mavromaras et al. (2010b) and Mavromaras and McGuinness (2012). Recall, the motivation for these studies was discussed (in a footnote) in Section 2.3.

There are the limitations that result from the empirical identification of over-education, which ultimately mean over-education research does not consider: the potential for individuals' innate abilities, quality of educational institution attended and qualification vintage to affect the human capital derived from education; whether the completion of multiple qualifications renders individuals over-educated; and, whether the human capital individuals derive from education is entirely relevant for their job, in terms of subject matter. There is also the lack of consistency in the method used to empirically identify over-education and, more importantly, little correlation between the three methods used. With regards to these methods, this study has argued that the JA method is to be preferred. There is the endogeneity bias and data limitations that mean most over-education wage penalty estimates cannot be interpreted as causal effects (i.e., no clear evidence on whether over-education has a causal effect on individuals' wages). There is the lack of evidence regarding the dynamics of over-education (e.g., individuals' entries, exits and durations over-educated, the effects by individuals' durations over-educated and the effects following individuals' exits from the state). And there is the focus on wages as the only (private) benefit to education investments, which means the literature has overlooked the role individuals' preferences may play in the incidence of over-education (i.e., the possibility some over-educated individuals have achieved their preferred outcome).

In outlining the key limitations of the over-education literature above, the intention was not to raise concerns over the validity of the research field—after all, given the importance of education and labour markets in modern societies, research into the utilisation of human capital is clearly warranted—but rather to highlight the areas in which advances are needed. Future contributions to the over-education literature should, therefore, aim to address one (or more) of these limitations.

Chapter 3

Research design

3.1 Introduction

As stated in Chapter 1, this study aims to investigate the existence of over-education in Australian labour markets. This chapter outlines the research undertaken to achieve that aim. In particular, Section 3.2 defines and discusses the research questions considered, and Section 3.3 describes the research method chosen to empirically test them (i.e., the data, sample examined, identification of over-education and data analyses performed). Remaining sections provide some further details. Section 3.4 describes the data used and discusses its suitability for examining over-education in Australia, while Section 3.5 describes the method used to identify over-education.

3.2 Research questions

To investigate the existence of over-education in Australian labour markets, this study considers the following research questions:

Do individuals who are identified as over-educated have human capital from education that is under-utilised in their current job?	[RQ1]
Are instances of over-education short-term labour market disequilibria that have no enduring effects?	[RQ2]
Are some individuals *voluntarily* over-educated (i.e., trading wages for non-pecuniary benefits of employment or improved working conditions)?	[RQ3]

RQ1 concerns the empirical identification of over-education; an empirical test of it therefore assesses the validity of the method used to identify over-education. The key assertion in over-education research is that individuals identified as over-educated are employed in jobs that do not fully utilise their human capital from education. Hence, determining whether empirical evidence supports this assertion is critical for establishing evidence of over-education in Australian labour markets. If evidence supports the assertion, then instances of over-education represent a source of potential gains for the economy—having over-educated individuals instead well-matched would result in higher wages for the individuals and then higher productivity levels, economic growth rates and

living standards.[1] If there is no evidence to support the assertion, then over-education may be merely an artificial phenomenon identified in data—a statistical artefact (McGuinness, 2006)—which is, most likely, the result of assumptions made regarding labour markets and individuals' accumulation of human capital from education (i.e., assuming jobs have productivity ceilings and that all individuals acquire the same human capital from a given qualification). Thus, individuals deemed over-educated are actually employed in jobs fully utilising their human capital from education and, therefore, instances of over-education do not represent a source of potential gains for the economy. Clearly, the two possibilities have distinct implications for government policy: the former suggests policy interventions that prevent and resolve instances of over-education would benefit the Australian economy, while the latter suggests no such policy interventions are necessary as labour markets are already effectively matching individuals and jobs (i.e., there are no individual-job mismatches leading to under-utilisation of the human capital individuals derive from education).

RQ2 concerns the possibility that over-education is a result of the inherently dynamic nature of labour markets. Modern labour markets are complex and continuously evolving because the labour supply and labour demand are always changing: there are always individuals investing in human capital, moving in and out of the labour force and moving between jobs, and firms are always creating, adjusting and discontinuing jobs. This may result in the individual-job mismatches that lead to over-education. But, with time, it may also lead to their resolution as either the individuals, in seeking to maximise their utility (or wages), move to jobs that fully utilise their human capital, or the firms, in seeking to maximise their profits, adjust jobs to fully utilise individuals' human capital. Hence, instances of over-education may be short-term disequilibria. If the time spent over-educated has no enduring effects (i.e., effects that occur even after individuals are no longer over-educated), over-education could then be regarded as merely a by-product of adjustment processes in dynamic and well-functioning labour markets. In this case, government policy interventions, if deemed necessary, should be aimed at increasing the speed of such adjustments in order to minimise the costs of over-education. If, however, over-education is persistent for individuals or has enduring effects, then it represents more serious labour market failures, and government policy interventions would be necessary to prevent and resolve such instances of over-education. The most likely enduring effect, which is considered in this study, is that time spent over-educated could lead to human capital depreciation. Such a link would mean the costs of over-education are greater than first thought as they not only arise while individuals are over-educated, but also in future employment where individuals appear well-matched (i.e., their labour productivity will be lower compared to if

[1] The realisation of such gains is, of course, predicated on the assumption that well-matched jobs either already exist or can be created for these individuals (and that they do not displace other well-matched individuals). This study does not consider whether this is actually the case, nor does it consider the means by which over-education can be prevented or resolved.

they had not previously been over-educated). And this would increase the importance of government policy interventions to prevent over-education.

RQ3 concerns the possibility that not all instances of over-education represent labour market failures. Given many attributes of jobs (e.g., work hours, job security and required effort), and not just the wage, can affect individuals' utility levels, it is possible some over-educated individuals have obtained jobs that maximise their (expected) utility levels and, therefore, achieved their preferred outcome. Such *voluntarily over-educated* individuals would trade wages for non-pecuniary benefits of employment or improved working conditions and, as a result, accept jobs for which they are over-educated. Since *voluntary over-education* arises due to individuals' preferences regarding job attributes, it would not be considered to be the result of labour market failures, at least from the perspective of individuals. Hence, no government policy interventions would be necessary to prevent or resolve such voluntary over-education, as doing so would result in welfare losses for individuals. From the perspective of societies, however, voluntary over-education is arguably still representative of labour market failures given it is associated with the under-utilisation of the human capital available in the workforce (i.e., an inefficient allocation of resources).

Each of these research questions addresses limitations and unresolved issues in the over-education literature. Recall from Chapter 2, existing research is weakened by disagreement and uncertainty regarding the theoretical basis for over-education and the method used to empirically identify it. As a result, the validity of over-education as a labour market phenomenon and the validity of its identification are unresolved issues. RQ1 has been formulated to consider such issues. Also, since the empirical test of RQ1 is based on an examination of the relationship between over-education and individuals' wages, it is intended to address the limitation that most over-education wage penalty estimates cannot be interpreted as causal effects. RQ2 responds to the lack of evidence on the dynamics of over-education and, in particular, it has been formulated to consider whether examining labour markets from a dynamic perspective, rather than the static approach typically used, diminishes the importance of over-education as labour market failures. RQ3 has been formulated to address the fact existing research has focused on wages as the only benefit to education investments and thereby overlooked the role individuals' preferences may play in over-education.

3.3 Research method – Empirical tests of the research questions

The research questions are empirically tested using the following research method. First, individual-level survey data are used to examine instances of over-education. In particular, the Household,

Income and Labour Dynamics in Australia (HILDA) Survey data for 2001 to 2008 period are used.[2] The HILDA Survey data was chosen because it is a longitudinal survey designed to contain a sample representative of the Australian population residing in private dwellings. The longitudinal, or panel, nature of the data enables examination of the behaviour of the same sample of individuals over time. This proves valuable for both descriptive analyses (e.g., analysis of the persistence of over-education) and multivariate analyses, where it allows for the controlling of unobserved individual heterogeneity. Meanwhile, the representativeness of the data enables inferences to be made regarding over-education throughout Australian labour markets. A further reason for choosing the HILDA Survey data is the rich array of information it contains on the past and present circumstances of individuals in the sample. This enables previously unexplored factors concerning over-education to be examined (e.g., the relationship between over-education and individuals' preferences) and means multivariate analyses can control for detailed sets of individual characteristics, which likely improves the accuracy of parameter estimates.[3]

The sample of individuals examined is then restricted. Throughout this study the focus is on employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.[4] The reasons for these sample restrictions are as follows. First, the sample is restricted to working-age, employed individuals because over-education directly relates to individuals' employment outcomes. Second, full-time students are omitted because their employment situation is likely quite different from that of other individuals. For many full-time students, employment may serve as merely a source of financial support and is subject to the varying time constraints associated with student schedules.[5] In addition, this group includes individuals yet to enter labour markets on a full-time and permanent basis (e.g., secondary school students). Any such individuals who are employed may be regarded as investing in their human capital (i.e., acquiring skills from work experience) to prepare for future careers, or as merely seeking a source of supplementary income. Regardless, full-time students will typically seek flexible jobs rather than jobs that fully utilise their human capital (or represent career jobs).[6] Finally, self-employed individuals are omitted because, similar to full-time students, their employment situation is distinct from other individuals. Their role

[2] The in-confidence unit record files from HILDA Survey data Waves 1 to 8 (Release 8.0), February 2010.
[3] The HILDA Survey data has one major limitation for the study of over-education: the absence of any information on the subject matter (or *field of education*) of individuals' qualifications—hence, its role in the incidence of over-education cannot be explored here. Note, however, this deficiency is to be rectified in Wave 12, where additional questions on skills, abilities and education are to be included. The HILDA Survey data is further discussed in Section 3.4.
[4] These sample restrictions are applied to each wave of the HILDA Survey data separately; hence, it is not the case that the conditions must hold across all eight waves for an individual to be included in the sample.
[5] That is, main availability for work being weekends and nights during semesters, unavailability during examination periods and then increased availability during semester breaks.
[6] Part-time students, however, are examined because they are likely to have entered labour markets on a full-time and permanent basis and be undertaking education to either facilitate a change in career or progress in their current career (i.e., increasing human capital (or *up-skilling*) to obtain a promotion). Such transitions may be important means for entering or exiting over-education; part-time students therefore remain in the sample.

as both employer and employee should result in more flexible jobs and no asymmetric information regarding the availability and use of human capital. As a result, self-employed individuals should be able to tailor their jobs to match their skills, and thereby ensure the greatest possible utilisation of their human capital.[7] This is clearly different to the employment circumstances of individuals working as employees.

Since it was established as the preferred method in Section 2.3, the JA method is used to identify instances of over-education. Recall, this involves comparing the level of education completed by each individual with the required education level of their current job. In this study, such comparisons are performed in terms of AQF qualification levels, rather than years of education, because using years of education does not account for the possibility that some of the time spent in education may be ineffective (e.g., course failures and repeated years at school) (Sloane et al., 1999). And the use of qualification levels arguably better reflects modern labour markets, whereby firms assess individuals on the basis of their qualifications completed, rather than years of education, and individuals invest in education in order to obtain such qualifications. The JA method requires an occupation classification that explicitly defines the required skill (or education) level of each job in the labour market being examined. For Australian labour markets, there are two such occupation classifications: the ABS (1997) Australian Standard Classification of Occupations (ASCO) and the ABS (2006) Australian and New Zealand Standard Classification of Occupations (ANZSCO). This study uses ANZSCO because it is the more recent classification and should, therefore, more accurately represent the jobs in Australian labour markets during the 2001–2008 period.[8] A further benefit in using the HILDA Survey data is that it provides ANZSCO information at the highly detailed 4-digit level.[9, 10] This means the main assumption associated with the JA method—that individuals classified to the same occupation category are employed in jobs with the same required

[7] This is predicated on two assumptions: first, self-employed individuals, in their role as employer, seek to maximise profits (i.e., fully utilise the available human capital); second, there is a demand for the resultant output.

[8] That is, given the time between the creation of ASCO and the period examined, there may be jobs in the 2001–2008 period that are not captured in ASCO (i.e., new jobs). Since job requirements are expected to change over time, most likely due to technological advances, ANZSCO is also likely to contain more accurate measures for required education levels. Preliminary analyses indeed indicate significant differences in estimates of over-education based on ASCO and ANZSCO; these results are not presented, but are available from the author on request. A further reason for choosing ANZSCO is that it defines skill levels in terms of those necessary to *perform* jobs, whereas ASCO refers to skills necessary to *obtain* jobs. Hence, the ANZSCO definition of skills is consistent with the desired interpretation of over-education (i.e., relates to the utilisation of skills in jobs), but this is not necessarily true for the ASCO definition. There are also empirical reasons for choosing ANZSCO. Specifically, for HILDA Survey data, it is recommended that occupations be examined using ANZSCO variables, as they likely have a lower coding error rate than ASCO variables (Watson and Summerfield, 2009). Furthermore, ANZSCO variables are available in all years (or waves) of the data, but ASCO variables are not.

[9] The 4-digit level ANZSCO information is only available in the in-confidence unit record files, not the confidentialised (general release) unit record files, of the HILDA Survey data. For details on the structure of ANZSCO see Appendix 3.1.

[10] To the author's knowledge, the HILDA Survey data is the only data source for Australia (suitable for studying over-education) that provides access to unit record data on individuals and has ANZSCO information at the 4-digit level. Alternative (publicly available) data sources only provide ANZSCO information at the 1-digit or 2-digit levels (e.g., the ABS Census of Population and Housing data (ABS, 2009)).

education level—can be made with greater confidence.[11] Hence, the disaggregated level to which occupations are reported in the HILDA Survey data helps to reduce the likelihood that identified instances of over-education arise due to measurement error (though this may be somewhat offset by the increased potential for measurement error that arises when occupations are coded to more highly disaggregated levels).[12]

Based on these data, the chosen sample and the instances of over-education identified, multivariate analyses are then used to perform the empirical tests for each research question. Specifically, regression models and semi-parametric matching methods are used to estimate certain effects (or parameters) of interest.[13] Details on these analyses are presented in sub-sections below. Multivariate analyses are used because they can control for potentially confounding factors (i.e., factors correlated with the variable of interest and the outcome being examined), which then isolates the effect a particular factor has on an outcome. Thus, the estimated effects more accurately represent *causal effects*, particularly when the analyses control for both observed and unobserved individual factors.[14] This approach to empirical testing is consistent with modern applied economics research (see, for example, Angrist and Pischke (2009)).

3.3.1 Over-education and under-utilised human capital

For RQ1, the focus is individuals' labour productivity achieved in the workplace, where evidence of reduced labour productivity is considered indicative of individuals having human capital that is under-utilised in the job. It is assumed wages reflect the labour productivity achieved, and hence the relationship between over-education and wages is examined.[15] Specifically, the effect over-education has on individuals' wages is estimated. Finding over-education has a negative, statistically significant effect—being over-educated *causes* an individual's wage to be lower than if they were instead well-matched—is, therefore, assumed to be evidence the individuals identified as over-educated do indeed have human capital from education that is being under-utilised. Thus, it would validate the method used to identify over-education. An important requirement for such a conclusion is that the estimates represent causal effects, not just conditional associations. But, as discussed in Section 2.4, estimating

---

[11] This is because the large number of occupation categories at the 4-digit level in ANZSCO—358 categories (compared to only 43 categories at the 2-digit level (see Appendix 3.1))—increases the likelihood that individuals in the same category perform jobs that are similar and, as a result, have the same required education level. In fact, according to ABS (2006), it is sufficient to use the 358 categories at the 4-digit level to examine occupations by skill level, because at this level virtually all jobs in the same category have the same required skill level.

[12] ANZSCO and the identification of over-education in this study are further discussed in Section 3.5.

[13] Also, at various stages, descriptive statistics (i.e., means, variances and correlations) are used in the empirical tests (e.g., the analysis of the persistence of over-education in Chapter 6).

[14] The definition and estimation of causal effects is discussed in Section 5.3.

[15] Assuming the wages paid to individuals reflect the value of their labour productivity achieved in the workplace is a standard assumption in modern economics research, and the basis of human capital theory (see, for example, Becker (1964), Mincer (1974), Rumberger (1987), Card (1999) and Dearden, Reed and Van Reenen (2006)).

such causal effects is not straightforward. The concern is that unobserved individual heterogeneity (or non-random selection into over-education) may lead to biased estimates. In response, a series of empirical estimators, some of which are designed to control for unobserved individual heterogeneity, are used to derive estimates of the over-education wage penalty. The robustness of the resultant estimates is then examined, with the aim being to establish over-education wage penalty estimates that can be interpreted as causal effects. Estimates are derived using the pooled OLS, fixed effects, first-differences, cross-sectional (or propensity score) matching, combined matching and regression, and difference-in-differences (or longitudinal) matching estimators. The use of these estimators is justified and explained in Chapter 5.

This approach has one key limitation: the effect of over-education may vary across individuals (e.g., by highest education level or duration over-educated) and the derived estimates represent merely the *average* effect among all over-educated individuals. Given such variation, some individuals would experience effects above the average and some below, and for some individuals (likely only a small number) there may be no discernible effect (i.e., some of the over-educated individuals may not incur a wage penalty). Hence, finding that there is, on average, a statistically significant over-education wage penalty does not necessarily guarantee that all the individuals identified as over-educated have under-utilised human capital. The approach is also reliant on the assumption that wages accurately reflect individuals' labour productivity achieved in the workplace and, as stated above, the ability to derive estimates that represent causal effects.

3.3.2 Dynamics of over-education

For RQ2, two sets of analyses are conducted. The first considers whether over-education is merely short-term disequilibria. As discussed in Sections 2.2 and 3.2, instances of over-education can be resolved in two ways: either individuals move to jobs that fully utilise their human capital or firms adjust jobs to fully utilise individuals' human capital. In determining whether over-education is temporary or persistent for individuals, it is therefore necessary to consider evidence of resolutions from both these sources. Since HILDA Survey data capture individuals' movements between jobs, evidence of resolutions from the first source is derived via examination of the changes in individuals' over-education statuses over time.[16] Hence, descriptive statistics on individuals' transitions to and from over-education and their durations over-educated are calculated. Such analyses, however, may not capture evidence of resolutions from the second source because individuals may not report

[16] HILDA Survey data actually capture individuals' changes in occupations, not jobs. But, since occupations are examined using ANZSCO at the 4-digit level, such changes are sufficient for examining resolutions of over-education because any job changes occurring within these occupation categories (which cannot be identified in the data) are unlikely to result in a change in over-education status (as, recall, virtually all jobs in the same occupation category of ANZSCO at the 4-digit level have the same required skill level (ABS, 2006)). HILDA Survey data on occupations is further discussed in Section 3.4.2.

adjustments to their jobs made by their employer as a change of occupation. That is, if the adjustments to individuals' jobs that lead to the full utilisation of their human capital cannot be identified in the data, then affected individual-job matches would appear unchanged and the individuals, erroneously, remain identified as over-educated. In order to ascertain whether such resolutions occur, the relationship between duration over-educated and individuals' wages is examined. Specifically, the over-education wage penalty is estimated by individuals' duration over-educated, and evidence the wage penalty disappears with time spent over-educated is considered indicative of over-education being resolved by firms adjusting jobs.[17] In analyses similar to the empirical test of RQ1, the fixed effects estimator is used in an attempt to derive estimates that can be interpreted as causal effects. Finding individuals are only temporarily identified as over-educated or the over-education wage penalty disappears over time is, therefore, assumed to be evidence that instances of over-education are short-term labour market disequilibria.

The second set of analyses for RQ2 considers whether over-education has enduring effects, and in particular, the possibility that over-education leads to human capital depreciation. Two empirical tests are performed. One examines whether being over-educated in the past increases the likelihood of being over-educated again in the future—that is, it tests for state dependence in over-education. Specifically, the effect that being over-educated in the previous year has on an individual's current likelihood of being over-educated is estimated, with evidence that it increases the likelihood considered indicative of state dependence in over-education. Since state dependence is defined in terms of a causal relationship, it is important the estimates represent causal effects. This requires an econometric model (and empirical estimators) capable of separately identifying the causal effects of prior over-education and (observable and unobservable) individual characteristics. Thus, dynamic panel probit models (and the Heckman (1981c), Orme (1997) and Wooldridge (2005) estimators) are used to derive the estimates. These are justified and explained in Chapter 6. The other empirical test examines the relationship between prior over-education and the wages of currently well-matched individuals. Specifically, the effect that being over-educated in the previous year has on their current wages is estimated, with evidence that it reduces wages considered indicative of individuals having reduced labour productivity due to human capital depreciation.[18] Analyses similar to the empirical test of RQ1 are again performed, with the difference-in-differences matching estimator used in an

[17] That is, since it is again assumed wages reflect the labour productivity achieved in the workplace, evidence the wage penalty disappears is considered indicative of individuals who appear to have been over-educated for extended periods no longer actually having under-utilised human capital—instead, their human capital is utilised to the same extent as (otherwise identical) well-matched individuals—and, presumably, these instances of over-education were resolved because firms adjusted the jobs to fully utilise their human capital.

[18] That is, since it is assumed wages reflect labour productivity and the sample is restricted to currently well-matched individuals, evidence the previously over-educated have reduced wages, compared to (otherwise identical) individuals who were well-matched, is considered indicative of these individuals having reduced labour productivity due to human capital depreciation (which presumably occurred while they were over-educated).

attempt to derive estimates that can be interpreted as causal effects. Finding state dependence in over-education or prior over-education reduces the wages of well-matched individuals is assumed to be evidence that over-education leads to human capital depreciation. And, therefore, it would be evidence that instances of over-education have enduring effects.[19]

These empirical tests of RQ2 have some limitations. First, the descriptive analyses may not capture all of the variation in over-education status over time as data on individuals' changes in occupations, rather than jobs, are examined. This may lead to the persistence in over-education being overstated. Second, evidence of state dependence in over-education is not necessarily indicative of human capital depreciation. Such evidence may instead be the result of discrimination or stigma effects associated with over-education (e.g., firms mistakenly believing previously over-educated individuals are less productive, which adversely affects the job offers received by these individuals) or the fact individuals face costs of exiting over-education that exceed the benefits (e.g., the cost of geographical relocation exceeding the expected wage increase). Third, similar to the empirical test of RQ1, *average* effects are estimated. This means, for example, that finding prior over-education has a negative, statistically significant effect on the wages of well-matched individuals does not necessarily guarantee that all the previously over-educated individuals experience such a wage reduction and, therefore, human capital depreciation. Finally, the analyses are reliant on the assumption that wages accurately reflect individuals' labour productivity achieved in the workplace, and the ability to derive estimates that represent causal effects.

3.3.3 Voluntary over-education

For RQ3, it is necessary to develop a method by which instances of voluntary over-education, if they exist, can be identified. This study proposes a distinction based on individuals' job satisfaction levels and their desire for a new job (or intentions to quit current job): it is assumed over-educated individuals who report being both highly satisfied with their job and highly unlikely to quit are *voluntarily over-educated*, while the remainder are *involuntarily over-educated*. The method and resultant estimates must then be validated. If valid, there should be evidence of trade-offs between wages and other job attributes among the individuals identified as voluntarily over-educated, and hence the

[19] Evidence of human capital depreciation and a disappearing wage penalty would also have implications for the identification and interpretation of over-education. If individuals' human capital depreciates while over-educated, then it is possible some individuals no longer possess the human capital from their education that is, or was previously, under-utilised in their current job; thus, they are incorrectly identified as over-educated and would, at least notionally, be better characterised as well-matched. Alternatively, if no human capital depreciation occurs and the wage penalty disappears with time spent over-educated, then it is possible some individuals are no longer under-utilising their human capital from education and would, as a result, also be better characterised as well-matched. Such evidence, therefore, would mean not all instances of over-education identified in data necessarily represent individuals with under-utilised human capital. But, to be clear, in the first instance previously over-educated individuals become 'well-matched' due to a loss of human capital, while in the second instance it is a greater utilisation of their human capital that leads to the change. Both cases would also then affect the costs of over-education.

relationship between voluntary over-education and job attributes is examined. Specifically, the differences in job attributes between voluntarily over-educated and well-matched individuals, and between the involuntarily over-educated and well-matched, are estimated. Finding that, compared to well-matched individuals, the voluntarily over-educated experience wage penalties and improvements in other job attributes and that the involuntarily over-educated incur wage penalties without such improvements is assumed to validate the identification of voluntary over-education. And, therefore, it would be evidence that some individuals are indeed voluntarily over-educated. A number of job attributes are considered, including: the wage; hours worked (and whether they are individuals' preferred hours); type of employment contract; job security; degree of flexibility to decide when to work (including whether the job has flexible start and finish times); degree of autonomy; and, degree of stressfulness. Since job attributes are treated as the dependent variables in the analyses and they vary between being measured as continuous, ordinal and binary, three different econometric models and estimators are used: the linear fixed effects, fixed effects ordered logit and fixed effects probit estimators.[20] Each uses a regression framework and the panel data to control for (observable and unobservable) individual heterogeneity, thereby ensuring the estimated differences in job attributes are not confounded by other factors. Use of such estimators is justified and explained in Chapter 7.

This empirical test has three limitations. First, it adopts a static approach to labour markets as it is assumed evidence of *contemporaneous* trade-offs—wage penalties and improvements in other job attributes that are experienced in the same time period—validates the existence of voluntary over-education. This is a simplification because, in reality, individuals are likely to make employment decisions with respect to maximising their *lifetime* utility levels. Second, choosing the levels of job satisfaction and intentions to quit that represent 'highly satisfied' and 'highly unlikely to quit' is somewhat arbitrary and, since different values will produce different estimates, this leads to some uncertainty regarding estimates of voluntary over-education. Third, and similar to the empirical test of RQ1, the estimated differences in job attributes are the *average* differences for all voluntarily over-educated (and all involuntarily over-educated) individuals. Hence, the derived evidence would not necessarily guarantee that all the individuals identified as voluntarily over-educated experience a wage penalty and improvements in other job attributes. The approach is also reliant on the assumption that these estimated differences are not confounded by other factors (i.e., reflect only the differences that arise from being voluntarily (and involuntarily) over-educated rather than well-matched).

[20] Fixed effects probit estimator approximated using random effects probit estimator with Mundlak-Chamberlain controls.

3.4 The HILDA Survey data

The HILDA Survey is a longitudinal household survey designed to contain a sample of individuals that is representative of the Australian population residing in private dwellings. Commencing its first wave in 2001, the HILDA Survey consisted of a large national probability sample of Australian households and personal interviews with all members aged 15 years and older in the selected households. The members of responding households in Wave 1 are the foundation of the panel tracked in future waves, with any extensions to the sample the result of changes to the composition of these original households. In Wave 1, interviews were conducted at 7,682 households, from which 13,969 individuals were successfully interviewed. Interviews occur between August and March each year on a roughly annual basis for each individual, and are typically administered on a face-to-face nature.[21] For further details on the design and progress of the HILDA Survey see Wooden and Watson (2007).

To ensure the HILDA Survey data are suitable for examining over-education in Australia, several factors must be considered. The first, given the method used to identify over-education, is the quality of the education and occupation data, and in particular, the extent of incomplete information (or item non-response) in variables for individuals' highest education level and ANZSCO occupation, and the identification of changes in (or the consistency of) these variables across waves. The accuracy of such data is critically important because any mismeasurement may lead to measurement error in the identification of over-education. Moreover, the accuracy of changes in these variables is important because this study examines the dynamics of over-education and also uses such dynamics, in multivariate analyses, to control for unobserved individual heterogeneity. The second factor is the representativeness of the sample in the HILDA Survey data, particularly with regard to individuals' education levels and occupations, as this study seeks to estimate the incidence and draw inferences concerning over-education throughout Australian labour markets. The final factor is sample attrition. Since this study frequently utilises the longitudinal nature of the data (i.e., focuses on samples of individuals who were respondents in numerous waves in order to observe dynamics), it is possible such sample attrition can bias results. That is, if the individuals who drop out of the sample are significantly different from those who remain, then results may not accurately reflect behaviour in the wider (or true) population. While there is a non-trivial amount of sample attrition in the HILDA Survey data, this does not necessarily mean results will be biased.[22] For biased results, the sample attrition must be non-random in nature and the characteristics associated with

[21] The vast majority of interviews occur between September and November each year.
[22] Comparing the number of initial respondents to the number of individuals who responded in each wave—commonly referred to as the *balanced panel*—provides an indication of the amount of sample attrition in the HILDA Survey data: there are 13,969 initial respondents and 8,034 individuals in the (W1–W8) balanced panel.

attrition must also be correlated with the instances of over-education (or, more generally, educational mismatches) (Fitzgerald, Gottschalk and Moffitt, 1998; Watson and Wooden, 2004).

The following sub-sections consider each of these factors. Sections 3.4.1 and 3.4.2 scrutinise the quality of education and occupation data and, in response to the data quality issues identified, they discuss the adjustments (or data cleaning) performed on the HILDA Survey data. By and large, the HILDA Survey data has very few data quality issues. One important issue, however, is the inconsistency of the ANZSCO occupation variables across waves (Watson and Summerfield, 2009). This must be addressed. Also, based on comparisons with other Australian data sources, these sections consider the representativeness of the sample in the HILDA Survey data. But, such analyses are only briefly discussed because, for both highest education level and ANZSCO occupation information, it is difficult to obtain data that can be used as an accurate benchmark (i.e., data regarded as a definitive representation of the Australian population). Section 3.4.3 then considers the potential for sample attrition in the HILDA Survey data to lead to biased results in this study.

3.4.1 Education data

The HILDA Survey data contain an extensive amount of information on each individual's education. When individuals first appear in the HILDA Survey they are asked a series of questions concerning their educational history. To begin, they are asked about their school education: whether still attending; age left; highest year completed; type attended; and, whether completed last year overseas. Following this, individuals are asked about their post-school education: whether ever enrolled in course for trade certificate, diploma, degree or other qualification; qualifications completed; country in which completed highest qualification; and, whether currently enrolled in a course for qualification.[23] In all subsequent waves, individuals are then asked about educational activities undertaken since their last interview: whether spent any time as a school student; whether left school; whether spent any time in a course for qualification; whether completed a course for qualification; and, whether currently enrolled in a course for qualification.[24] Based on the responses to these questions, variables for the highest education level of each individual are derived, whereby the classification of qualifications into an ordinal scale (necessary when constructing such measures) is based on the ABS (2001) Australian Standard Classification of Education (ASCED).[25]

---

[23] The series of questions on school and post-school education can be found in Section C of the Wave 1 Person Questionnaire (for initial respondents to the HILDA Survey) or Section A of the New Person Questionnaires for Wave 2 to Wave 8 (for individuals who first respond after Wave 1). For a copy of these questionnaires see the HILDA Survey webpage concerning the survey instruments: http://www.melbourneinstitute.com/hilda/doc/questionnaires/default.html.

[24] The series of questions on educational activities since last interview can be found in Section A of the Continuing Person Questionnaires for Wave 2 to Wave 8. As above, copies of these questionnaires can be found on the HILDA Survey webpage concerning the survey instruments.

[25] Ordinality is not always guaranteed in the ASCED at Broad Levels 5 (Certificate Level) and 6 (Secondary Education). Specifically, in instances where an individual has both a Certificate Level qualification and Secondary Education it may not

In the HILDA Survey data, the highest education level of each respondent is almost always identified. There are, in fact, only two sources of incomplete information: individuals who have an 'undetermined' highest education level (due to item non-response during the survey); and, individuals who report having completed a Certificate Level qualification, but do not provide information on its particular level (i.e., whether it is a Certificate I, II, III or IV). As Table 3.1 indicates, there are very few individuals with such incomplete information. Nevertheless, these cases must be addressed. Since over-education status cannot be identified without highest education level information, the individuals with an 'undetermined' highest education level are omitted from the sample examined. Given such a small number of observations are lost, this sample restriction is unlikely to impart bias on results of subsequent analyses. For individuals with 'Certificate – not further defined', it is assumed that their highest education level is a Certificate IV. This is because most are older individuals who likely obtained their qualification (a trade certificate) prior to the introduction of the current classification system (i.e., Certificates I-IV) and their qualifications are likely equivalent to a Certificate IV (i.e., the current certification required to be a qualified tradesman is a Certificate IV).[26]

Table 3.1: Number of cases of incomplete information in highest education level variables—All individuals (N)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Undetermined	6	7	6	4	6	5	6	5
Certificate – not further defined	86	74	77	77	77	77	70	67
N	13,969	13,041	12,728	12,408	12,759	12,905	12,789	12,785

SOURCE: Author's calculations using HILDA Survey data (Release 8.0); specifically, tabulation of the derived variables for highest education level ('_edhigh') in each year.

With variables for individuals' highest education level in each wave established, the next concern is the changes in (or consistency of) these variables across the eight waves of data (for each individual). In particular, it is necessary to determine which changes appear real and which, if any, appear to result from inconsistency (or measurement error) in the variables. This is done by examining the changes across consecutive waves (i.e., W1–W2, W2–W3, W3–W4, etc.), along with the reported completions of school and post-school qualifications (since each individual's last interview). Table 3.2 reports the results. The figures indicate the vast majority of individuals do not

always be the case that the qualification at the Certificate Level is to be regarded as the highest. For these cases, the ABS has developed a Decision Table for determining the individual's highest education level (see ABS (2002)). This Decision Table consists of a cross-tabulation of the qualifications at these two Broad Levels and an indication of which is to be regarded the highest. According to supporting documentation (see, for example, Summerfield (2010)), the derived variables for highest education level in the HILDA Survey data utilise this Decision Table where necessary.

[26] Note, however, instances of over-education cannot be identified simply using the HILDA Survey derived variables for highest education level as, in order to enable comparisons to the required education levels defined in ANZSCO, variables with (slightly) more disaggregated categories are needed (e.g., it is necessary to make the distinction between Advanced Diploma and Diploma qualifications and between each of the four levels of Certificate qualifications). This is done using the information on all the qualifications completed by each individual.

complete any study for a qualification between interviews and, as a result, there is no change in their highest education level. But there is a small fraction, roughly 3 to 4 per cent, whose highest education level does change. Given the nature of education, whereby an individual cannot lose a previously obtained qualification, only increases in highest education levels should be observed; decreases are viewed as the result of likely measurement error. Closer inspection of Table 3.2, in fact, highlights three potential sources of measurement error in the highest education level variables.

Table 3.2: Number of changes in highest education level by reported completion of study—Individuals who respond in consecutive waves (N)

	W1–W2	W2–W3	W3–W4	W4–W5	W5–W6	W6–W7	W7–W8
Increase in highest education	395	436	414	481	424	467	464
Completed study	*355*	*377*	*368*	*432*	*384*	*408*	*402*
No completed study	*40*	*59*	*46*	*49*	*40*	*59*	*62*
No change in highest education	11,596	11,046	10,861	10,856	11,282	11,353	11,321
Completed study	*327*	*315*	*319*	*342*	*282*	*287*	*287*
No completed study	*11,269*	*10,731*	*10,542*	*10,514*	*11,000*	*11,066*	*11,034*
Decrease in highest education	2	0	1	0	2	2	0
N	11,993	11,482	11,276	11,337	11,708	11,822	11,785

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

The first is the individuals whose highest education levels decrease between waves. For this small number of individuals, the decreases appear to represent measurement error and, therefore, some (minor) alterations are made to the highest education level variables.[27] The second is individuals whose highest education level increases, but who do not report the completion of a school year or a post-school qualification since their last interview.[28] Further examination of the data indicates the measurement error actually lies in the identifier for having completed study between interviews. Specifically, these individuals were enrolled in the final years of secondary school at the time of the two interviews (i.e., Year 11 and Year 12) and had a highest education level of Year 11 in the first interview and Year 12 in the second interview, but then did not report completing any study between the interviews. This appears to simply be the result of recall bias. The identified increases in highest education level are, therefore, considered accurate. The third possible source of measurement error is individuals who report completing study since their last interview, but whose highest education level does not increase.[29] Examination of the data reveals these individuals all complete qualifications at or below the level of their highest education in the first interview. Their highest education levels, therefore, should remain unchanged, and so these cases do not represent measurement error.

[27] This is based on an examination of the highest education level of these individuals in each wave. For details on the analysis see Appendix 3.2: Table A3.2.1 presents their highest education levels in each wave, while Table A3.2.2 specifies the alterations made to the highest education level variables.
[28] This is the case for roughly 10 per cent of the individuals whose highest education level increases between waves.
[29] This is the case for roughly 50 per cent of the individuals who report completing a school or post-school qualification since their last interview.

Overall, the highest education level variables in the HILDA Survey data appear to be of the quality needed to study over-education: few observations are lost due to incomplete information and, following a small number of (minor) alterations to ensure their consistency across waves, the changes in these variables appear to represent real instances of individuals increasing their highest education levels between interviews.

Assessing the representativeness of the sample in the HILDA Survey data with regards to levels of education—whether the distribution of individuals across highest education levels adequately represents that in the Australian population—is a difficult task. This is because there is no data regarded as a definitive representation of highest education levels in the Australian population. Several ABS surveys attempt to derive such data, but are typically subject to limitations. This study has considered three of them. First, the ABS Census of Population and Housing, which is conducted every five years in Australia and has the unique benefit of being drawn from the entire population, produce data with significant numbers of individuals having incomplete information regarding their education.[30] Second, the ABS Survey of Education and Work (SEW), which is an annual supplement to the (monthly) ABS Labour Force Survey, tend to produce data with samples that, according to the supporting documentation, have over-sampled individuals whose highest education is Year 12 or a Bachelor Degree and under-sampled individuals with a Graduate Diploma, Graduate Certificate or Certificate III–IV highest education.[31] Finally, the ABS Survey of Education and Training (SET), which is conducted every four years in Australia, produce data that, given the lack of significant limitations, appear the most likely to accurately represent the Australian population.[32] The SET data are, therefore, used to benchmark the HILDA Survey data.[33, 34]

Results of the benchmarking exercise, which are presented in Table A3.3.1 (the comparison to SET data for 2001 and 2005) and Table A3.3.2 (the comparison to SEW data for each year in the 2001 to 2008) in Appendix 3.3, suggest the HILDA Survey data may contain higher proportions of highly educated individuals—those with highest education levels of a Postgraduate Degree, Graduate Diploma or Certificate, Bachelor Degree, Advanced Diploma, Diploma or Certificate III–IV—than

[30] For example, in the 2001 and 2006 Census data, almost 25 per cent of the individuals with a non-school qualification had 'not stated' or 'inadequately described' responses regarding their highest non-school qualification (ABS Cat. no. 2068.0). Further details on the Census of Population and Housing can be found in ABS (2009).
[31] For further details on the SEW and the Labour Force Survey see, for example, ABS (2007a) and ABS (2007b).
[32] For further details on the SET see, for example, ABS (2010).
[33] Despite the limitations, the SEW data are also used in the benchmarking exercise, mainly because they provide annual estimates that can be compared to each wave of the HILDA Survey data.
[34] Differences in survey methodologies are, of course, likely to affect the comparability of these data. In particular, the survey questions on education (e.g., asking individuals all qualifications completed vs. highest qualification, and use of an open-ended question vs. a question with a list of acceptable answers) and the manner in which the questionnaires are administered (e.g., self-administered vs. telephone interviews vs. face-to-face interviews) can lead to differences. Different weighting regimes may also affect the comparability of these data. For the data considered here, there are indeed differences in methodology (e.g., the HILDA Survey is administered by face-to-face interviews, while the Census is self-administered, the SEW is (mostly) administered by telephone interviews and the SET is administered by face-to-face interviews.

actually exist in the Australian population. And then, at the other end of the spectrum, it may also contain lower proportions of less educated individuals—those with only secondary school education. But, since the representativeness of the benchmarking data cannot be guaranteed, this is not conclusive evidence. Instead, it is interpreted as a potential caveat for this study: if the HILDA Survey data have indeed over-sampled highly educated individuals, then this study may over-estimate the incidence of over-education in Australian labour markets.[35]

3.4.2 Occupation data

One of the main sections in the HILDA Survey questionnaires concerns the current employment situation of each individual, where information on work hours, number of jobs, tenure, industry, contracts and entitlements, sector and job satisfaction levels are all collected.[36] Also contained in this section is the survey question regarding the kind of work individuals do in their current job. Specifically, the question asks for the occupation title and a description of the main tasks and duties performed.[37] Based on this information, variables are derived that classify the jobs of each individual to occupations as defined in the ABS (2006) ANZSCO. These ANZSCO occupation variables are available at the 1-digit, 2-digit and 4-digit levels.[38]

Similar to individuals' highest education levels, the ANZSCO occupation of each individual is almost always identified in the HILDA Survey data. There are three types of incomplete information in the ANZSCO occupation variables: individuals with 'not stated' (due to item non-response during the survey); individuals who responded 'don't know' to the survey question; and, individuals who did not supply sufficient information for their job to be classified to an ANZSCO occupation category at the 4-digit level. In the latter case, individuals are classified to a less detailed level of ANZSCO and distinguished as 'not further defined (n.f.d)'.[39] As Table 3.3 indicates, there are few individuals with such incomplete information, and they are almost all 'n.f.d' cases. Since over-education cannot be identified without the ANZSCO occupation information, the individuals with

---

[35] This is based on the assumption that highly educated individuals are more likely to be over-educated; thus, data that have over-sampled the highly educated are likely to produce upwardly biased estimates of the incidence of over-education.

[36] The series of questions on current employment can be found in Section E of the Wave 1 Person Questionnaire and Section C of the Continuing (or New) Person Questionnaires for Wave 2 to Wave 8. For a copy of these questionnaires see the HILDA Survey webpage concerning the survey instruments:
http://www.melbourneinstitute.com/hilda/doc/questionnaires/default.html

[37] This is survey question E13 in the Wave 1 Person Questionnaire and C11 in the Continuing (or New) Person Questionnaires for Wave 2 to Wave 8. These survey questions are only asked of individuals who have previously indicated they are employed, and for individuals with multiple jobs they refer only to their *main* job—where their main job is deemed to be the one for which they usually receive the most income each week.

[38] Recall, the 4-digit level is the most detailed (or disaggregated) level of information available in the data and such ANZSCO occupation variables at the 4-digit level are only available in the in-confidence unit record files of the HILDA Survey data. For details on the structure of ANZSCO see Appendix 3.1.

[39] For example, if an individual can be classified to the 'Specialist Managers' category at the 2-digit level but cannot be classified to an occupation category at the 4-digit level, then in the ANZSCO occupation variables at the 4-digit level they are classified as 'Specialist Managers – n.f.d'. Refer to Table A3.1.2 in Appendix 3.1 to see the presence of such 'n.f.d' categories in the ANZSCO classification at the 2-digit level.

'not stated' and 'don't know' are omitted from the sample examined. This is unlikely to impart bias on future results as so few observations are lost. For the 'n.f.d' cases, the available information on occupations (at the higher level of ANZSCO) can be used to identify required education levels for these individuals, though it may require some assumptions.[40] These are considered later as 'n.f.d' cases can be resolved in the subsequent cleaning of the ANZSCO occupation variables.

Table 3.3: Number of cases of incomplete information in ANZSCO occupation variables—All employed individuals (N)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Not stated	0	0	0	1	0	0	0	0
Don't know	11	0	0	0	0	0	0	1
Not further defined (n.f.d) (4-digit level)	61	27	21	20	33	34	1	56
N	8,525	8,088	7,991	7,822	8,247	8,357	8,342	8,369

SOURCE: Author's calculations using HILDA Survey data (Release 8.0); specifically, tabulation of the derived variables for ANZSCO occupation ('_jbmo06') in each year.

As already mentioned, the consistency of the ANZSCO occupation variables across waves is a concern (Watson and Summerfield, 2009). This is clearly observed when the changes in ANZSCO occupation variables across consecutive waves are examined. Panel A of Table 3.4 reports the results of such an examination. The figures indicate that, based on ANZSCO at the 4-digit level, approximately 40 per cent of individuals employed in consecutive years change their occupation. These estimates, given a priori expectations and knowledge of labour market mobility, appear excessive. To confirm this, estimates based on alternative methods for identifying occupation change are examined. In the HILDA Survey data, changes in occupation can be identified using three other methods. There is the survey question that directly asks individuals whether they have changed occupation since their last interview.[41] And, somewhat less reliably, changes can also be identified using information on individuals' years of tenure in their current occupation and whether they report being promoted at work during the last year.[42] For such measures a change in occupation is assumed if tenure is less than one year and if individuals report being promoted at work. These are, however, imperfect measures for occupation change. First, the measure based on tenure may not capture changes where individuals move to occupations they previously held and where their tenure is greater

[40] In the ANZSCO documentation, 'n.f.d' categories are referred to as supplementary categories to enable the processing of inadequately described responses in survey data and, as a result, do not have a defined required education level. Hence, it may be necessary to assume required education levels for the 'n.f.d' categories observed in the data based on the required education levels of related occupations.

[41] This is survey question C12a in the Continuing Person Questionnaires for Wave 2 to Wave 8.

[42] Information on tenure in current occupation is based on survey question E14 in the Wave 1 Person Questionnaire (or C12 in the New Person Questionnaires for Wave 2 to Wave 8) and C12b in the Continuing Person Questionnaires for Wave 2 to Wave 8. Information on being promoted at work during the past 12 months is based on survey question B16 in the Self Completion Questionnaires for Wave 2 to Wave 8 (question number changes from Wave 5 onwards, as follows: B14 in Wave 5; B24 in Wave 6; B23 in Wave 7; B20 in Wave 8).

than one year. Second, the measure based on promotion may not capture changes where individuals move to a new employer (i.e., use of the phrase 'promoted at work' in the survey question may be interpreted by some individuals as meaning they must have remained with the same employer).

Table 3.4: Number of changes in occupation, based on various methods of identification—Individuals employed in consecutive waves (N)

Changes in occupation	W1–W2	W2–W3	W3–W4	W4–W5	W5–W6	W6–W7	W7–W8
A. Comparison of ANZSCO occupation variables							
ANZSCO 1-digit level	1,782	1,719	1,656	1,610	1,681	1,886	1,872
	(27.0%)	(26.5%)	(25.6%)	(24.2%)	(23.8%)	(26.0%)	(26.4%)
ANZSCO 2-digit level	2,424	2,318	2,278	2,225	2,299	2,586	2,574
	(36.7%)	(36.2%)	(35.3%)	(33.4%)	(32.8%)	(35.8%)	(36.2%)
ANZSCO 4-digit level	2,983	2,803	2,777	2,751	2,868	3,134	3,123
	(45.0%)	(43.4%)	(43.0%)	(41.0%)	(41.5%)	(44.0%)	(43.6%)
B. Alternative methods to identify changes							
Self-reported change in occupation since last interview							
Yes	1,123	1,085	1,084	1,176	1,260	1,361	1,399
	(17.5%)	(16.9%)	(16.3%)	(17.7%)	(17.8%)	(18.9%)	(19.7%)
Tenure in current occupation							
Less than 1 year	776	792	798	870	881	1,000	976
	(12.3%)	(12.1%)	(11.9%)	(13.1%)	(12.1%)	(13.7%)	(13.5%)
Promoted at work in past year							
Yes	552	585	608	585	625	657	663
	(8.8%)	(9.0%)	(9.1%)	(8.8%)	(8.6%)	(9.0%)	(8.7%)
N	6,744	6,602	6,538	6,649	7,016	7,084	7,142

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses use balanced panels of respondents for the pairs of consecutive waves of data.
Weighted proportions are reported in parentheses; these are weighted using longitudinal population weights to make them representative of the Australian population.
Since the Self Completion Questionnaire (SCQ) of the HILDA Survey has a non-trivial amount of non-response (ranging from 6.5 to 11.0 per cent of respondents per wave), the figures for promotion may under-estimate the actual proportion of individuals who are promoted at work each year in Australian labour markets.

Panel B of Table 3.4 presents the estimates based on these alternative methods. The figures suggest 9 to 20 per cent of the individuals employed in consecutive years are changing occupation. These are significantly lower than the estimates derived using the ANZSCO occupation variables. While the alternative methods are expected to under-estimate occupational change, it is unlikely to be to an extent that would explain such differences.[43] The evidence in Table 3.4, therefore, supports the conclusion that the comparison of ANZSCO occupation variables across waves overstates occupational change in Australian labour markets. This has two important implications. The first is that there must be measurement error associated with the derivation of the ANZSCO occupation variables. For instance, there may be inconsistency in the descriptions of job tasks and duties

[43] The self-reported measure may under-estimate occupational change due to recall bias. That is, the need for individuals to accurately recall past events may result in measurement error that leads the variable to under-estimate occupational change. The accuracy of the measure depends on individuals recalling: whether they recently changed job; whether it occurred in the reference period between interviews; and, whether it was significant enough to constitute a change in occupation (in terms of tasks, duties and responsibilities). Given all the information required, there is scope for cases of change to go unreported.

reported by respondents (i.e., individuals describing the same job differently across waves), inconsistency in the coding of these descriptions to ANZSCO occupation categories (i.e., HILDA data entry persons coding the same descriptions of job tasks and duties differently across waves), or both concurrently. Thus, in many instances where occupation change is identified (i.e., ANZSCO occupation categories between two consecutive years are different for an individual) it is actually the case that the ANZSCO occupation category is inaccurately defined in (at least) one of the adjacent years. The second implication is that a more accurate method is required for identifying changes in occupation within the HILDA Survey data.

The inconsistency in the ANZSCO occupation variables is addressed as follows: 'definite' changes in occupation are first identified, then the ANZSCO occupation variables are adjusted such that they are consistent across waves for individuals who are not deemed to have had a 'definite' change in occupation. An individual is assumed to have had a 'definite' change in occupation between two waves if the ANZSCO occupation variables identify a change and (in the second wave) they report either: (i) a change in occupation (i.e., a self-reported change); (ii) current occupation tenure of less than one year; or, (iii) being promoted at work in the past year. Essentially, information from all four of the above methods is combined. The aim is to reduce the chance that measurement error in any one of them will affect the identification of changes and increase the likelihood of identifying all 'definite' changes in occupation that occur in the HILDA Survey data.

Based on the definition above, the identification of 'definite' changes in occupation is straightforward. Adjusting the ANZSCO occupation variables to ensure their consistency across waves, however, is not. The aim is to ensure that if an individual is employed across consecutive waves and does not have a 'definite' change in occupation, then their ANZSCO occupation category remains unchanged. Instances where it does change are assumed to represent measurement error. For each individual with an instance of measurement error, the sequence of ANZSCO occupation categories reported immediately before and after this measurement error is identified, whereby such a sequence is only broken if the individual has a 'definite' change in occupation, is not employed or is a non-respondent in a particular wave. This sequence of ANZSCO occupation categories is, therefore, assumed to be referring to a single occupation.[44] Panel A of Table 3.5 presents the number of individuals in each wave whose ANZSCO occupation category is part of such a sequence and which is, therefore, potentially mismeasured. The figures indicate a widespread problem; up to 45 per cent of individuals in each wave have occupations that are potentially misclassified to ANZSCO at the 4-digit level. This, however, over-states the extent of the problem as these sequences also contain the accurately measured occupation categories. That is, if an individual employed in all eight wave of data

[44] For individuals with multiple instances of measurement error, multiple sequences of ANZSCO occupation categories are identified and then considered separately.

without any 'definite' changes in occupation has one instance of measurement error, then, given the above approach, their ANZSCO occupation category appears as potentially mismeasured in all eight waves even though it may be accurately measured in seven of them.

In order to achieve consistency, the ANZSCO occupation category in each wave of such sequences is assumed to be equal to the category that appears most frequently within the sequence. For cases where more than one most frequent category is observed (i.e., categories appearing the same number of times and are the most frequent within the sequence), the ANZSCO occupation category in each wave of the sequence is assumed to be equal to the first (chronologically) of the multiple most frequent categories. By definition, this assumption also applies to cases where there is no most frequent category within the sequence (i.e., different categories observed in each wave of the sequence).[45] This process is undertaken with the ANZSCO occupation variables at the 4-digit level, then, based on these revised (now consistent) 4-digit level variables, the 2-digit level and 1-digit level ANZSCO occupation variables are derived. Panel B in Table 3.5 reports the number of ANZSCO occupation categories that were revised as part of this process. The figures indicate that, at the 4-digit level, roughly 10 to 20 per cent of the ANZSCO occupation categories of employed individuals in each wave required revision.

Table 3.5: Number of revisions to ANZSCO occupation variables—All employed individuals (N)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
A. Accuracy of ANZSCO occupation variables								
ANZSCO 1-digit level								
Potentially mismeasured	1,957	2,324	2,449	2,464	2,522	2,520	2,387	1,999
Accurately measured	6,568	5,764	5,542	5,358	5,725	5,837	5,955	6,370
ANZSCO 2-digit level								
Potentially mismeasured	2,454	2,984	3,138	3,152	3,229	3,253	3,059	2,480
Accurately measured	6,071	5,104	4,853	4,670	5,018	5,104	5,283	5,889
ANZSCO 4-digit level								
Potentially mismeasured	2,949	3,611	3,736	3,772	3,871	3,897	3,688	2,971
Accurately measured	5,576	4,477	4,255	4,050	4,376	4,460	4,654	5,398
B. Number of revisions made to ANZSCO occupation variables								
ANZSCO 1-digit level	501	841	770	781	719	782	734	875
ANZSCO 2-digit level	701	1,196	1,088	1,136	1,032	1,136	1,069	1,223
ANZSCO 4-digit level	904	1,528	1,391	1,467	1,327	1,462	1,369	1,584
N	8,525	8,088	7,991	7,822	8,247	8,357	8,342	8,369

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: 'Potentially mismeasured' indicates an individual's current ANZSCO occupation category is part of a sequence that contains an instance of measurement error, while 'accurately measured' indicates that this is not the case.

[45] This assumption is, of course, somewhat arbitrary. But, it is justified if individuals are believed likely to exert the most effort in describing their occupation in the first interview in which they report having it. If, however, inconsistency in the ANZSCO occupation variables results from inconsistency in the coding of these variables, rather than inconsistency in individuals' descriptions, then this decision is more difficult to justify.

Given the revised ANZSCO occupation variables (and the previously defined method for identifying 'definite' changes in occupation), revised estimates for the rate of occupational change in Australian labour markets are derived. Table 3.6 presents the results. The figures indicate that, based on ANZSCO at the 4-digit level, approximately 15 to 18 per cent of individuals employed in consecutive years change their occupation. Such estimates are significantly lower than the initial estimates reported in panel A of Table 3.4, and are in accordance with a priori expectations regarding labour market mobility. Hence, it is assumed that these identified cases of occupational change adequately represent those occurring in Australian labour markets.

Table 3.6: Number of 'definite' changes in occupation—Individuals employed in consecutive waves (N)

Changes in occupation	W1–W2	W2–W3	W3–W4	W4–W5	W5–W6	W6–W7	W7–W8
Comparison of *revised* ANZSCO occupation variables							
ANZSCO 1-digit level	714	703	709	719	749	927	893
	(11.2%)	(11.0%)	(11.0%)	(10.8%)	(10.5%)	(12.4%)	(12.3%)
ANZSCO 2-digit level	919	889	903	915	945	1,165	1,153
	(14.4%)	(14.1%)	(13.8%)	(13.9%)	(13.1%)	(15.7%)	(16.2%)
ANZSCO 4-digit level	1,042	978	1,001	1,039	1,080	1,286	1,287
	(16.4%)	(15.4%)	(15.3%)	(15.8%)	(14.9%)	(17.5%)	(18.0%)
N	6,744	6,602	6,538	6,649	7,016	7,084	7,142

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Analyses use balanced panels of respondents for the pairs of consecutive waves of data.
Weighted proportions are reported in parentheses; these are weighted using longitudinal population weights to make them representative of the Australian population.

The derivation of the revised ANZSCO occupation variables also provides the opportunity to minimise the amount of incomplete information (i.e., 'n.f.d' cases) in these variables. In particular, if an individual has 'n.f.d' in a particular wave but valid information in a wave (or waves) immediately before or after it and no 'definite' change in occupation between them (i.e., an 'n.f.d' case within a sequence of ANZSCO occupation categories describing a single occupation), then this information (i.e., the ANZSCO occupation category at the 4-digit level) is used to resolve the 'n.f.d' case. Such adjustments are included in the number of revisions reported in panel B of Table 3.5, while panel A of Table 3.7 reports the number unable to be resolved. These figures, when compared to those in Table 3.3, indicate a significant reduction in the amount of incomplete information. It is then necessary to deal with the few remaining cases. Recall, the individuals with 'not stated' and 'don't know' are omitted from the analyses. For the remaining 'n.f.d' cases, it is necessary to assume required education levels. In some instances, the assumptions are straightforward as the 4-digit categories to which the 2-digit 'n.f.d' category is related all have the same required education level. For example, if an individual's occupation is classified as 'Education Professionals' at the 2-digit level but unable to be classified at the 4-digit level—such that in the 4-digit ANZSCO occupation variables they are classified as 'Education Professionals – n.f.d'—then regardless of which 4-digit

level category they should actually be classified to (e.g., 'Primary School Teachers', 'Secondary School Teachers', 'University Lecturers and Tutors') the required education level of their occupation is the same—specifically, a Bachelor Degree or higher. Thus, their required education level can be assigned without introducing any measurement error. Panel B of Table 3.7 shows that this occurs for the vast majority of 'n.f.d' cases in the revised ANZSCO occupation variables. For a small number of individuals, however, there is a degree of uncertainty and the assumed required education levels have the potential to introduce some measurement error.[46]

Table 3.7: Number of cases of incomplete information in revised ANZSCO occupation variables—All employed individuals (N)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
A. Incomplete information								
Not stated	0	0	0	1	0	0	0	0
Don't know	11	0	0	0	0	0	0	1
Not further defined (n.f.d) (4-digit level)	26	11	7	4	8	12	0	25
B. Consequences for identification of required education levels (ANZSCO at 4-digit level)								
Assumed – No measurement error	23	11	7	3	5	10	0	25
Assumed – Potential measurement error	3	0	0	1	3	2	0	0
Observations dropped	11	0	0	1	0	0	0	1
N	8,525	8,088	7,991	7,822	8,247	8,357	8,342	8,369

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

Overall, the ANZSCO occupation information in the HILDA Survey data appear to be of the quality needed to study over-education. Similar to the highest education level variables, few observations are lost due to incomplete information. A pertinent concern, however, is the inconsistency of the ANZSCO occupation variables across waves. But, following revisions aimed at identifying 'definite' changes in occupation and ensuring consistency in the classification of individuals' occupations, the changes in the revised ANZSCO occupation variables appear to represent real instances of individuals changing occupation between interviews.

Once again, assessing the representativeness of the sample in the HILDA Survey data is a difficult task. That is, determining whether the distribution of individuals across ANZSCO occupations adequately represents that in the Australian population is difficult because there is no data regarded as providing a definitive representation of the Australian population. This study has considered data from two other Australian sources: the ABS Labour Force Survey (LFS) and the ABS Census of Population and Housing.[47, 48] Results of the benchmarking exercise, which are

[46] For such assumptions the aim is to provide a broad approximation, and so the required education levels of all the 4-digit categories to which the particular 2-digit 'n.f.d' category relates are used to create a range for the required education level.
[47] As with the benchmarking of the highest education level information, differences in survey methodologies are likely to affect the comparability of these data. The main concern with Census data is that the self-administered nature of the survey may result in less reliable occupation information, because the failure to respond (or failure to respond in sufficient detail)

presented in Table A3.4.1 (the comparison to LFS data for each year in the 2001 to 2008 period) and Table A3.4.2 (the comparison to Census data for 2006) in Appendix 3.4, suggest the HILDA Survey data may contain higher proportions of individuals employed in Professional (and, to a lesser extent, Community and Personal Service Worker) occupations than actually exist in the Australian population. And then, as a counterbalancing effect, it may contain lower proportions of individuals employed as Technicians and Trades Workers, Clerical and Administrative Workers, Sales Workers and Labourers. But, since the representativeness of the benchmarking data cannot be guaranteed, this is not conclusive evidence. Instead, it is once again interpreted as a potential caveat for this study: if the HILDA Survey data have indeed over-sampled individuals employed as Professionals, then this study may under-estimate the incidence of over-education in Australian labour markets. This is because, based on the method used to identify over-education, such Professionals are assumed to have required education levels of a Bachelor Degree or higher and, therefore, cannot be deemed over-educated. The method used to identify over-education in this study is discussed below in Section 3.5.

3.4.3 Sample attrition

As with all longitudinal surveys, the HILDA Survey data is subject to sample attrition: individuals who were interviewed at one wave ceasing to participate in the survey at a later wave because they either died, moved overseas, could not be located or refused to continue participating (Watson and Wooden, 2004). Such sample attrition, however, is not necessarily permanent as individuals can rejoin the survey following a wave (or waves) of non-response (i.e., the attrition is non-monotonic). In the HILDA Survey data, the (unadjusted) wave-to-wave attrition rates—proportion of Wave *t–1* respondents who did not respond in Wave *t*—have declined from 14.1 per cent in Wave 2 to 7.9 per cent in Wave 8.[49] And, compared to the 13,969 initial respondents, there are 8,034 individuals in the (W1–W8) balanced panel. There is clearly a non-trivial amount of sample attrition in the data.

---

to occupation questions will result in more 'inadequately described' and 'n.f.d' cases than in data based on surveys administered by face-to-face interviews (such as the HILDAS Survey). Inspection of Census data, such as presented in Table A3.4.2 in Appendix 3.4, confirms that this is indeed case. On the other hand, a concern with LFS data is that the estimates used for comparison are annual averages derived using the published quarterly data, unlike the HILDA Survey data which refer to a particular point in time (roughly the September to November period of each year) (also, LFS data are derived (mostly) from telephone interviews). Recall, ABS (2009) provides details on the ABS Census of Population and Housing, while ABS (2007a) and ABS (2007b) provide details on the ABS LFS.

[48] A further issue is the level of ANZSCO (i.e., 1-digit, 2-digit or 4-digit level) at which the benchmarking is performed. Since this study uses ANZSCO at the 4-digit level to identify over-education, it may be expected that the benchmarking be performed at the 4-digit level. At this level, however, there are more than 350 occupation categories and, given the HILDA Survey sample size is small relative to this level of detail, it appears unreasonable to expect such a sample to perfectly represent Australian labour markets. Hence, the benchmarking analyses are instead performed with respect to ANZSCO at the 1-digit level (using the LFS data) and 2-digit level (using the Census data).

[49] These rates include individuals who died, moved overseas and could not be located as non-respondents, and they equate to the following loss of observations: 1,976 in Wave 2; 1,559 in Wave 3; 1,452 in Wave 4; 1,071 in Wave 5; 1,051 in Wave 6; 1,083 in Wave 7; and, 1,004 in Wave 8. For further details on sample attrition in the HILDA Survey data see, for example, Watson and Wooden (2004, 2009, 2011).

But, as previously stated, it does not necessarily mean the results in this study of over-education will be biased. For biased results, the sample attrition must be non-random and the characteristics associated with attrition must also be correlated with over-education (Fitzgerald et al., 1998; Watson and Wooden, 2004). This is considered below.

Evidence suggests sample attrition in the HILDA Survey data is indeed non-random; for instance, the probability of non-response is higher among individuals who are young (aged between 15 and 24 years), less educated (without a post-school qualification), migrants (from English speaking or non-English speaking background) or Aboriginal or Torres Strait Islander, single or in a de facto relationship, have poor health, had a problem or cooperated poorly in the previous interview, and are more mobile (i.e., those who have lived in numerous homes in the past ten years, reported a high likelihood of moving in the next 12 months (during the previous interview) and who actually moved homes between the two interviews) (Watson and Wooden, 2004; Watson and Wooden, 2009).[50] Thus, the key question that remains is whether the characteristics associated with attrition are also correlated with over-education. Or, more specifically, whether instances of over-education (and educational mismatch, in general) are correlated with future non-response. This is empirically tested using multivariate analyses. In particular, a series of probit models for the likelihood of non-response are estimated, whereby controls for educational mismatch status and individual, household and interview characteristics are included.[51]

Two-year panels of data are constructed using the adjacent waves of the HILDA Survey data (i.e., W1–W2, W2–W3, W3–W4, etc), and probit models for the probability of non-response in the later wave given the individual responded in the initial wave are then estimated. The controls in these models refer to individuals' characteristics in the initial wave. Specifically, three specifications are estimated: (I) contains only the controls for educational mismatch status; (II) adds a basic set of controls for individual characteristics (e.g., gender, age, ethnicity, marital status and education); and, (III) adds a more extensive set of controls for individual, household and interview characteristics (e.g., whether moved house between interviews at $t-1$ and t, income, area of residence and problems encountered during interview at $t-1$).[52] The models are estimated separately for each of the (seven) two-year panels of data, and then for all these data pooled together. For the latter models, year dummies (corresponding to the year at $t-1$) are included to account for possible time effects, and robust (panel-corrected) standard errors are calculated to account for the potential serial correlation in the error terms of individuals with multiple observations in these data. Ultimately, since the

[50] This evidence is also drawn from the analysis of non-response below (i.e., the estimated coefficients on controls for individual, household and interview characteristics derived from the subsequent probit models).
[51] These are similar to the models estimated in Fitzgerald et al. (1998) and Watson and Wooden (2009).
[52] The controls in Specification (III) closely resemble those included in the logistic attrition models used to create the longitudinal responding person population weights in the HILDA Survey data (see, for example, Watson (2004)). Except, of course, the controls for educational mismatch, which are those derived in this study (as described in Section 3.5).

estimated coefficients represent the conditional effects that each characteristic has on the likelihood of non-response, the focus is on the coefficients for educational mismatch. Finding statistically significant coefficient estimates is, therefore, assumed to be evidence that instances of educational mismatch are correlated with future non-response. Table 3.8 presents results for the models estimated using the pooled data.[53]

Table 3.8: Probit model estimates for probability of non-response at *t* given individual responded at *t-1*—Attrition from two-year panels across W1–W8

Dependent variable: Non-response $_t$	(I)		(II)		(III)	
	Coeff.	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Educational mismatch at t-1 (base: Not employed)						
Over-educated	0.052**	(0.020)	0.079**	(0.023)	-0.060	(0.409)
Under-educated	0.080**	(0.023)	0.029	(0.026)	-0.131	(0.410)
Well-matched	-0.063**	(0.014)	0.027	(0.016)	-0.098	(0.409)
Controls for basic individual characteristics	No		Yes		Yes	
Controls for extensive individual, household and interview characteristics	No		No		Yes	
N	90,599		90,599		90,599	
(No. non-respondents at *t*)	(9,196)		(9,196)		(9,196)	
Pseudo R^2	0.0096		0.0542		0.1120	
Log likelihood	-29,465.29		-28,136.76		-26,418.24	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.
Controls for basic individual characteristics: gender; age; ethnicity; English language proficiency; marital status; number of children; highest education; self-assessed health status; long-term health condition; life satisfaction.
Controls for extensive individual, household and interview characteristics: moved house between interviews at *t-1* and *t*; employment status and usual hours worked per week; personal income; rest of household (HH) income; income support recipient; relationship in HH; likelihood of moving home in next 12 months; number of homes during past ten years; area of residence; remoteness of area; SEIFA 2001 index of disadvantage; dwelling type; housing tenure; HH type; number of persons in HH; number of bedrooms per person in HH; number of adults in HH; number of children in HH; length of Person Questionnaire interview; did not complete Self-Completion Questionnaire; reference person in HH; number of calls to HH for contact; incomplete survey response in HH; interviewer observations regarding interview situation (e.g., problems encountered).
All models also contain year dummies (i.e., identifiers for year which corresponds to period *t-1*).
'Pseudo R^2' values equal 1 minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept term (Fitzgerald et al., 1998).
Complete sets of results are not presented, but are available from the author on request.

Based on the results from specification (I), educational mismatch appears correlated with future non-response; compared to being not employed, being over-educated or under-educated at *t–1* appears to increase the likelihood of non-response at *t*. Meanwhile, being well-matched has the opposite effect. With the inclusion of basic individual characteristics in specification (II) the effects of being under-educated and well-matched are weakened, and are no longer statistically significant.

[53] Results for the models estimated separately for each of the (seven) two-year panels are qualitatively similar to those for the pooled data; hence, they are not presented. However, such results, along with the complete set of results for models in Table 3.8, are available from the author on request.

Over-education continues to have an effect (and it has actually been strengthened). But, in the results from specification (III), the effect of being over-educated is no longer statistically significant (and has actually changed to a negative effect). The results indicate, therefore, that once an extensive set of individual, household and interview characteristics are controlled for the instances of educational mismatch are not correlated with future non-response. Thus, educational mismatches exhibited significant conditional associations in results of specifications (I) and (II) because they were acting as proxies for characteristics that are correlated with non-response and not present in these models.[54, 55]

Since evidence suggests over-education (or educational mismatch, more generally) is not correlated with sample attrition, it appears unlikely that such sample attrition will be a source of bias in results derived in this study. Nevertheless, all subsequent analyses seek to avoid such bias. Specifically, descriptive analyses use the population weights in the HILDA Survey data—the cross-sectional responding person population weights are used in static analyses (i.e., those examining each wave separately) and the longitudinal responding person population weights are used in dynamic analyses (i.e., those examining individuals across multiple waves)—in attempts to ensure estimates are representative of the Australian population.[56] On the other hand, multivariate analyses control for extensive sets of individual characteristics, such as those in the above non-response models, in attempts to ensure estimated effects are not confounded by any effects associated with sample attrition—that is, to ensure estimated effects represent the causal effect a particular factor has on an outcome.[57] Further, it has also been concluded that in multivariate analyses attrition is unlikely to impart significant bias on the estimated parameters of key factors, rather it is likely to affect the intercept terms of such models (Fitzgerald et al., 1998; Watson and Wooden, 2004). The estimated effects of over-education derived in this study therefore appear unlikely to be biased by the sample attrition in the HILDA Survey data.

---

[54] Since the balanced panel is often examined in this study, similar analyses are performed considering attrition from the balanced panel. The results are presented in Table A3.5.1 in Appendix 3.5, and are qualitatively similar to those above: educational mismatch does not appear correlated with attrition from the balanced panel in the HILDA Survey data.

[55] Also, given the importance of the dynamics of over-education in this study, similar analyses are performed considering whether changes in educational mismatch states (or their source: a change of occupation or increase in highest education)) are correlated with non-response. Results are in Table A3.5.2 in Appendix 3.5 and, once again, indicate no such correlation.

[56] This, of course, is entirely dependent on the quality of the weights used. A particular concern for the longitudinal population weights is the low explanatory power of the logistic regression models used to model attrition (Watson, 2004). These weights may, therefore, have little actual affect on estimates derived from descriptive analyses (Fitzgerald et al., 1998; Wooden and Watson, 2007). But, it can be argued that, since these models (with extensive sets of characteristics) explain little of the observed variation in probabilities of attrition, then such attrition may be predominantly random in nature and therefore unlikely to bias results (Fitzgerald et al., 1998; Watson and Wooden, 2004; Watson and Wooden, 2009).

[57] Population weights are not used in the multivariate analyses because such analyses are more concerned with explaining the behaviour of individuals—isolating the effect a particular factor has on individuals (i.e., identifying causal, or at least conditional, effects)—and less concerned with whether these data are representative of a certain population. This behavioural (or structural) approach to multivariate modelling means it is not necessary for any sample or population weights to be used (Cameron and Trivedi, 2005).

3.5 Method used to identify over-education

As stated in Section 3.3, this study uses the JA method and, in particular, ANZSCO to estimate the required education levels of jobs and, as a result, identify instances of over-education in the HILDA Survey data. ANZSCO was developed by the ABS, Statistics New Zealand and the Australian Government Department of Employment and Workplace Relations (DEWR) to enable the consistent analysis of information and statistics on occupations (ABS, 2006). It is a skill-based classification that refers to all jobs in the Australian and New Zealand labour markets. The classification is based on the definition of *jobs* and *occupations*: a *job* is assumed to be a set of tasks carried out by an individual for a particular employer (who then pays the individual a wage in return), while an *occupation* is a set of jobs that require a similar set of tasks be carried out. ANZSCO first groups jobs into occupations, then organises occupations into increasingly larger groups based on the similarity of tasks, which is judged in terms of the level and specialisation of skill needed to complete the tasks. The resultant structure of ANZSCO has five hierarchical levels (each denoted by codes with a certain number of digits): Major groups (1-digit codes); Sub-major groups (2-digit codes); Minor groups (3-digit codes); Unit groups (4-digit codes); and, Occupations (6-digit codes). The Occupations are the most detailed level of the classification, while the Major groups are the broadest. Specifically, there are 998 Occupations and 8 Major groups.[58]

ANZSCO also defines skill levels required to perform each occupation in the classification.[59] In particular, it defines five skill levels and assigns one (or more) of them to each group (at all five hierarchical levels) in the classification, thereby representing the skill level required to perform the occupations in that group.[60] The five skill levels are quantified in terms of levels of formal education and training (i.e., AQF qualification levels) and years of previous experience in a related occupation (i.e., relevant experience) that may substitute for this formal education.[61] Table 3.9 presents the skill levels defined in ANZSCO. These skill levels are used to estimate the required education levels of jobs in Australian labour markets. See ABS (2006) for the skill level assigned to each group at the Major, Sub-major, Minor, Unit and Occupations levels.

---

[58] The aggregation is as follows: the 998 Occupations are aggregated to form 358 Unit groups, which are aggregated to form 97 Minor groups, which are aggregated to form 43 Sub-major groups, which are aggregated to form the 8 Major groups. For further details on the structure of ANZSCO see Appendix 3.1.

[59] *Skill* is assumed to be the ability to competently perform the tasks of an occupation, whereby the skill level of an occupation is a function of its range and complexity of tasks—the greater the range and complexity of tasks the greater the skill level required for the occupation (ABS, 2006).

[60] According to ABS (2006), the skill levels specified for each occupation in ANZSCO are the result of consultations with employers, industry training bodies, professional organisations and others, whereby the aim was to maximise the likelihood that such information is an accurate and meaningful depiction of jobs in Australian labour markets.

[61] For some occupations, ANZSCO also states that on-the-job training may be an additional requirement. But, in these cases, the required amounts of on-the-job training are not specified and, as a result, such information is not incorporated into the identification of over-education in this study.

Table 3.9: ANZSCO skill levels and corresponding formal education levels and years of relevant experience

Skill level	Formal education (AQF qualification) level	Years of relevant experience to substitute for formal education
1	Bachelor Degree or higher qualification	At least 5 years
2	Diploma, Associate Degree or Advanced Diploma	At least 3 years
3	Certificate III (and two years on-the-job training) or Certificate IV	At least 3 years
4	Certificate II or Certificate III	At least 1 year
5	Compulsory Secondary School or Certificate I	-

SOURCE: ABS (2006) ANZSCO, First Edition (Cat. no. 1220.0), pages 7-8.

NOTE: For skill level 5, the ANZSCO documentation does not define 'Compulsory Secondary School' (i.e., the school year or level of school to which it refers). It is also stated that no formal education or on-the-job training may be required to perform some occupations with this required skill level.

Given its aim of grouping together jobs and occupations with similar tasks and required skill levels, ANZSCO is ideal for estimating the required education levels of jobs. Especially since the skill levels are explicitly defined as attributes of occupations—the skill level required to competently perform the tasks of the occupation—and not as a reflection of the skill levels of the individuals observed in each occupation. ANZSCO also satisfies the suitability criteria stated in Section 2.3. Specifically, since it was developed by the ABS, Statistics New Zealand and DEWR and in consultation with external stakeholders, it is reasonable to assume that acceptable levels of objectivity and precision were maintained during its construction. As it was released in 2006, and presumably constructed in the few years prior, its classification of occupations and definition of required skill levels should correspond closely to jobs in Australian labour markets during the 2001–2008 period examined. It also contains information at highly detailed (or disaggregated) levels (i.e., skill levels of categories at the 4-digit and 6-digit levels). In fact, according to ABS (2006), it is sufficient to use the 358 Unit groups (the 4-digit level) to examine occupations by skill level, because at this level virtually all jobs in the same group have the same required skill level. Since the HILDA Survey data contain information at this 4-digit level, the main assumption associated with the JA method—that individuals classified to the same occupation category are employed in jobs with the same required education level—can be made with reasonable confidence. And, as considered in Section 3.4.2 (and following some revisions), it appears individuals' jobs have been accurately coded to ANZSCO in the HILDA Survey data. Using ANZSCO and its required skill levels (at the 4-digit level) to estimate the required education levels of jobs in Australian labour markets during the 2001–2008 period, therefore, appears entirely reasonable.

There are, however, some issues that must be addressed in using ANZSCO to identify instances of over-education. The first, as indicated in Table 3.9, is that each of the five skill levels defined in ANZSCO is measured in terms of a range of formal education levels, rather than a single education level. Hence, required education ranges are defined for jobs (i.e., a minimum and a

maximum required education level is defined for each job).[62] As a result, this study is using broad approximations for the required education levels of jobs. While this may lead to under-estimation of the incidence of over-education (as the broad approximations make it less likely for individuals to be identified as over-educated), it means the instances of over-education that are identified are more likely to be valid (i.e., represent individuals whose human capital from education is not fully utilised). The approach is also realistic given ANZSCO defines skill levels for occupations, rather than jobs, and since there is likely to be some variation in the required education levels of jobs aggregated to the same occupation. A further issue concerns skill level 3. ANZSCO defines occupations with a skill level 3 as requiring either: a Certificate III and two years of on-the-job training, or a Certificate IV. Since individuals' years of on-the-job training completed in their current job cannot be reliably measured in the HILDA Survey data, this study assumes the Certificate IV is the sole required education level associated with skill level 3. Finally, for skill level 5, 'Compulsory Secondary School' is defined as the minimum required education level. ANZSCO, however, does not state the school year or level of school education to which this refers (e.g., Year 9 or Year 10). Based on the regulated minimum school leaving age in Australia, it is assumed this refers to Year 10.[63]

Given the required education ranges derived from ANZSCO, over-education is identified as follows. An individual is deemed *over-educated* if their education (or highest AQF qualification) level exceeds the maximum required education level considered necessary to perform their current job.[64] Such instances of over-education are interpreted as follows. Since they have human capital from education that is under-utilised in their current job and firms pay no wage premium for unused human capital, over-educated individuals do not receive the full benefits to their investments in education, at least in terms of increased earnings. These benefits are diminished but not necessarily zero, because without their education (or highest qualification) the over-educated individuals may have been unable to obtain and perform the job. Zero benefits are, of course, possible and would arise if individuals' second-highest qualification (or no qualification) would have been sufficient for

---

[62] This is amplified by the fact some occupation categories in ANZSCO have multiple skill levels—representing a range of required skill levels—assigned to them (e.g., 6121 *Real Estate Sales Agents* has required skill level 2–3). See Table A3.6.1 in Appendix 3.6 for the required education ranges defined in ANZSCO (for occupation categories at the 4-digit level).

[63] In the period examined, most Australian States and Territories had a legislated minimum school leaving age of 16 years. Most individuals turn 16 years of age during the completion of Year 10 at secondary school and, therefore, this study assumes the completion of 'Compulsory Secondary School' roughly corresponds to the completion of Year 10.

[64] As discussed in Section 2.2, individuals employed in high-skilled jobs (i.e., senior management and professional jobs) cannot be deemed over-educated because the maximum required education level defined for such jobs is a Postgraduate Degree (which cannot be exceeded). This accounts for the fact high-skilled jobs are unlikely to have productivity ceilings (i.e., the jobs are likely flexible and the individuals in them likely to have autonomy in designing them, such that additional human capital always leads to increased output). It is also consistent with the skill levels defined in ANZSCO, whereby senior management and professional jobs are classified to occupations with skill level 1: requiring a Bachelor Degree or higher qualification to be competently performed. In ANZSCO, such jobs are contained within the occupation categories: *Chief Executives, General Managers and Legislators*; *Farmers and Farm Managers*; *Specialist Managers*; *Arts and Media Professionals*; *Business, Human Resource and Marketing Professionals*; *Design, Engineering, Science and Transport Professionals*; *Education Professionals*; *Health Professionals*; *ICT Professionals*; and, *Legal, Social and Welfare Professionals*.

the job (i.e., the individuals are receiving no (work-related) benefit to their highest qualification). Similarly, diminished benefits do not necessarily imply over-educated individuals have over-invested in education. Over-investment concerns comparisons of costs and benefits—individuals have, arguably, over-invested when the costs of the under-utilised education exceed its benefits—but the study of over-education merely considers benefits of education. Ultimately, over-educated individuals could have obtained and performed their current job with the completion of less education (or a lower level qualification), which presumably would have been less costly to complete.

As in the over-education literature, a similar approach could then be used to identify instances of under-education—an individual whose education level is less than the minimum required education level of their job could be deemed under-educated. Such under-education, however, has limited relevance. Since the required education levels used in its identification are estimating the required human capital levels of jobs, under-education is typically interpreted as instances in which individuals have insufficient human capital to perform their current job. But, such an interpretation is inaccurate. While it indeed identifies instances where individuals' human capital from education is insufficient to perform their job, this does not necessarily mean their entire stock of human capital is insufficient. This is because, as previously discussed, individuals can derive the human capital necessary to perform their job from various sources other than education, most notably work experience and on-the-job training. Hence, the above method overlooks the potential in modern labour markets to substitute work experience for formal education.[65] To produce a more meaningful counterpart to over-education—under-education that indeed represents individuals with insufficient human capital to perform their job—this study attempts to account for such a substitution.

An approximation of the substitutability of work experience for formal education in Australian labour markets can be derived from the ANZSCO skill levels. In particular, as presented in Table 3.9, each of these skill levels has been quantified in terms of a formal education level and the years of previous experience in a related occupation (i.e., relevant experience) that may substitute for the formal education. Such information is used in the identification of under-education in this study. And each individual's years of relevant experience are measured using the HILDA Survey derived variables for tenure in current occupation.[66] Ultimately, an individual is deemed *under-educated* if their education (or highest AQF qualification) level is less than the minimum required education level considered necessary to perform their current job *and* their years of relevant experience are less than

[65] This potential substitution is particularly relevant for examinations of the general working-age population—as in this study—because it contains older cohorts of individuals who are more likely to have accumulated human capital from work experience (and on-the-job training) rather than formal education.
[66] These derived variables ('_jbocct') capture the total number of years individuals have worked in their current occupation, and not necessarily on a continuous basis.

those considered necessary to substitute for this formal education.[67, 68] Such under-educated individuals are interpreted as not having the human capital required to competently perform all the tasks of their current job (or perform all the tasks to the desired standard), and consequently they would not have obtained the job had an individual with the (higher) required education level or relevant experience applied for it. Hence, instances of under-education can be regarded as evidence of skill shortages in specific labour markets. The above approach, of course, has several limitations. Most importantly, the substitutability of experience for education defined in the ANZSCO skill levels is a coarse approximation of an undoubtedly complex relationship, whereby the extent to which experience can substitute for education in a job is likely to vary across firms, industries and individuals' attained education levels.[69] The extent to which years of employment impart human capital is also likely to vary across firms, industries and individuals (i.e., occupation tenure is an imperfect measure of the experience (or human capital) derived from previous employment). Nevertheless, compared to the standard approach in the literature, it better reflects the reality of modern labour markets. Recall, the focus on over-education means most analyses in this study are based on the comparison of over-educated and well-matched individuals, with the under-educated ultimately excluded from consideration. As a result, the approach also intends to reduce the number of individuals erroneously identified as under-educated and, therefore, improve the accuracy with which well-matched individuals are identified.

Collectively the states of over-education and under-education are often referred to as *educational mismatches* in this study. And, as previously discussed, the third possible state for an employed individual is to be well-matched. An individual is deemed *well-matched* if they are neither over-educated nor under-educated. Or, more specifically, if their education (or highest AQF qualification) level is within the required education range considered necessary to perform their current job *or* their years of relevant experience are equal to (or greater than) those considered necessary to substitute for such formal education. Well-matched individuals are therefore interpreted

[67] For some occupation categories with skill level 5, ANZSCO states no formal education or on-the-job training is required. Hence, it is assumed individuals employed in such occupations cannot be under-educated. For the required education ranges and the necessary years of relevant experience used to identify under-education see Table A3.6.1 in Appendix 3.6.
[68] To the author's knowledge, no previous studies have taken such an approach to identifying under-education. Notice, however, that the approach results in asymmetry in the definitions of over-education and under-education: under-education takes into account individuals' relevant work experience, while over-education does not (and, as such, the definition of over-education remains consistent with that used throughout the literature). Estimates of under-education based on the standard approach are presented (in Table 4.2 in Chapter 4) and discussed later.
[69] For instance, for a job that requires a Diploma (i.e., skill level 2 in ANZSCO), an individual who completed Year 10 may require three years of relevant experience to be able to competently perform the job, while an individual who completed a Certificate IV may require only one year of relevant experience. The coarse ANZSCO approximation fails to account for such variability and, as a result, would erroneously classify the second individual as under-educated. This approach, therefore, may not capture all the trade-offs between experience and education occurring in Australian labour markets.

as having human capital, at least in terms of formal education and work experience, which approximately matches that required for their current job.[70]

As discussed in Section 2.3 and above, the method used in this study to empirically identify instances of over-education is based on several key assumptions. The first is that levels of education (or AQF qualifications) can be used to quantify the required human capital levels of jobs—leading to the definition of a required education level for each job. The second is that individuals who complete the same level of education (or AQF qualification) acquire the same level of human capital from it. Additional assumptions then arise in using the JA method, specifically ANZSCO, to estimate the required education levels of jobs. It is assumed ANZSCO (and the required skill levels it defines) accurately represents the jobs in Australian labour markets during the 2001–2008 period examined. It is assumed all jobs within the same Unit group (or 4-digit level occupation category) in ANZSCO are sufficiently similar so as to have the same required education level (or range). It is assumed the required education levels (or ranges) of jobs remain fixed throughout the eight-year period examined. And, in the HILDA Survey data examined, it is assumed the information on individuals' highest education (or AQF qualification) level and current occupation (classified to ANZSCO at the 4-digit level) is accurate. This method and its assumptions lead to some important limitations for the study. First, this study does not consider the potential for individuals' innate abilities, quality of educational institution attended and qualification vintage (or year completed) to affect the human capital derived from education. Second, it does not consider whether the completion of multiple qualifications (at the same ASCED level or different levels) may render an individual over-educated. Third, it does not consider whether the human capital individuals derive from education is entirely relevant for their job, in terms of subject matter (or field of study). The ramifications of the first limitation are, by far, the most serious: if the human capital derived from education varies by individuals' innate abilities, quality of educational institution attended or qualification vintage, then the identification method (and all subsequent research) may be invalid (i.e., the individuals deemed over-educated may not actually have human capital from education that is under-utilised in their current job). The other limitations—along with the use of required education *ranges* to identify over-education—merely mean this study may under-estimate the extent to which human capital from education is under-utilised in Australian labour markets.

[70] Ideally, well-matched individuals would be interpreted as fully utilising their human capital from education in their job. Recall, however, the use of required education ranges may lead to under-estimation of the incidence of over-education (i.e., some well-matched individuals may actually have human capital from education under-utilised in their current job). This would occur, for example, if the required education level of their job is actually equal to the *minimum* required education level in the range used and it is less than their attained education level. But, based on the required education ranges used (as presented in Table 3.6.1 in Appendix 3.6) and the distribution of occupations across these required education ranges in Australia (as later observed in Table 4.2), the scope for such situations appears relatively small.

Chapter 4

Incidence of over-education in Australia

4.1 Introduction

Based on the identification method discussed in Section 3.5, this chapter presents estimates of the incidence of over-education in Australian labour markets during the 2001–2008 period. It, along with the empirical test in Chapter 5, aims to determine whether there is evidence of over-education in Australian labour markets. Section 4.2 presents the estimates of over-education, and highlights how accounting for the potential substitution of work experience for formal education affects the identification of under-educated and well-matched individuals. Section 4.3 compares the estimates with those from previous studies examining Australian labour markets, with such comparisons serving as a preliminary means for validating the estimates derived in this study. Section 4.4 concludes the chapter.

4.2 Estimated incidence of over-education

Table 4.1 presents the estimated incidence of over-education in Australia for the 2001–2008 period. In particular, the sample of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed are classified into three groups: over-educated, under-educated and well-matched. Recall, Section 3.5 discussed the identification and interpretation of each group.[1] The figures in Table 4.1 indicate that approximately 20 per cent of the individuals are over-educated in each year, while roughly 10 per cent are under-educated and 70 per cent well-matched. For the

[1] Descriptive statistics on the measures used to identify the groups—highest education level, labour market experience and required education level of job (and years of relevant experience that may substitute for such education)—are presented in Appendix 4.1. Specifically, Table A4.1.1 presents the distribution of individuals by highest education (or AQF qualification) level for each year from 2001 to 2008. Results indicate the proportions of working-age employees with tertiary or vocational qualifications increased over time, while the proportion that did not complete Year 12 declined. Table A4.1.2 presents the total labour market experience, occupation tenure and employer tenure among individuals, where, in accordance with Becker (1964), total experience is considered indicative of *general human capital* and occupation tenure and employer tenure are indicative of *specific human capital*. Results indicate the vast majority of the sample has been employed for ten years or more, while roughly 8 per cent are in their first two years of working. Around one-third has worked in the same occupation for ten years or more, while a further one-third has done so for less than two years: this points to significant mobility between occupations. The mobility between employers appears even greater, with roughly 42 per cent having worked for their current employer for less than two years. Finally, Table A4.1.3 presents the distribution of individuals by required education level of job, and defines the years of relevant experience necessary to substitute for such formal education. Results indicate a slow shift towards more highly skilled jobs in Australian labour markets, though the evidence does not provide a complete picture of all jobs on offer (as it does not consider jobs of self-employed individuals, full-time students and employed individuals aged 65 years or over (and any less than 15 years of age), and it does not consider vacant jobs).

eight-year period, the estimates also indicate small increases in the incidences of educational mismatch: the proportion over-educated rose from 18.7 per cent in 2001 to 21.6 per cent in 2008, while the proportion under-educated rose from 9.2 to 10.4 per cent.[2]

Table 4.1: Incidence of over-education by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)	Overall (W1-W8)
Over-educated	18.7	19.7	20.8	20.4	20.6	21.6	21.3	21.6	20.7
Under-educated	9.2	10.6	10.1	10.8	10.8	10.9	10.9	10.4	10.5
Well-matched	72.0	69.7	69.2	68.8	68.6	67.5	67.9	68.1	68.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279	6,355	48,855

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Figures are proportions that sum to 100.0 for each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

As discussed in Section 3.5, unlike previous studies, the identification of under-education in this study accounts for human capital derived from labour market experience. The aim is to account for the potential in modern labour markets to substitute experience for education. The effect of this adaptation is observed by comparing the estimates in Table 4.1 with estimates derived using the standard approach in the literature, which are presented in Table 4.2. As expected, failure to account for labour market experience results in significantly more individuals being deemed under-educated: the standard approach estimates roughly 40 per cent of individuals are under-educated in each year, with the counterbalancing effect being that significantly fewer individuals are classified well-matched. Recall, this adaptation does not affect the identification of over-education.

Table 4.2: Incidence of over-education (using method that does not account for relevant work experience) by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

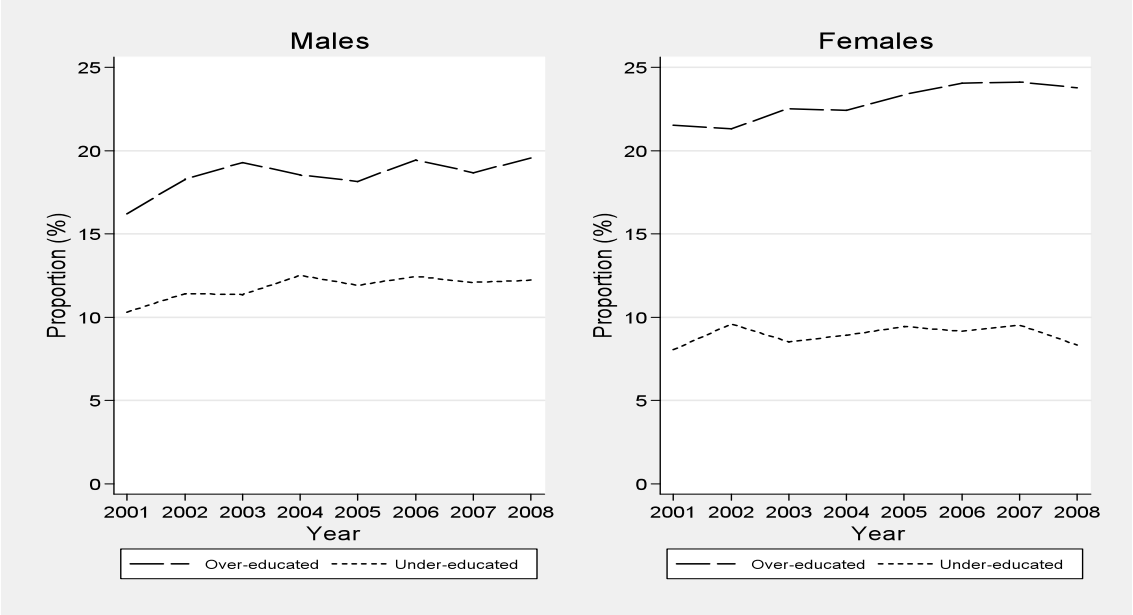
	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)	Overall (W1-W8)
Over-educated	18.7	19.7	20.8	20.4	20.6	21.6	21.3	21.6	20.7
Under-educated	40.3	40.5	38.9	37.8	37.2	36.9	36.8	34.9	37.8
Well-matched	40.9	39.8	40.3	41.9	42.2	41.5	41.9	43.5	41.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279	6,355	48,855

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Figures are proportions that sum to 100.0 for each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

[2] The reason for such increases is not considered here; this is a suitable avenue for future research, especially since 2001–2008 is generally regarded as a period in which labour market conditions in Australia improved. Of course, the increased incidence of over-education may result from assuming the required education levels of jobs remain fixed over the period (i.e., if they increased, then some of the individuals deemed over-educated in the later years may actually be well-matched).

Consistent with other labour market phenomena, particularly those concerning employment, it is common for over-education research to consider males and females separately. For instance, it has been argued that females are more likely to be over-educated due to the spatial constraints faced by married women (i.e., job search geographically constrained to where their husband is employed) or the adverse effects of career interruptions (e.g., time out of the labour force to have children) (Frank, 1978; McGoldrick and Robst, 1996). Figure 4.1 therefore presents the estimated incidence of over-education (and under-education) by gender.[3] The results indicate over-education is indeed more prevalent among females, ranging between roughly 21 and 24 per cent over the period, compared with 16 to 20 per cent for males. Similar to the overall results, the incidence of over-education increases slowly over time for both groups, though slightly more so for males. Meanwhile, under-education is more prevalent among males, with roughly 10 to 12 per cent of males under-educated, compared with 8 to 10 per cent for females. Its incidence is also rising over time for males, but is relatively unchanged for females. A further result, which is best observed in the accompanying table (Table A4.2.1 in Appendix 4.2), is that the proportion well-matched tends to be (somewhat) higher among males than females.[4]

Figure 4.1: Incidence of over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed



SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Figures are proportions and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

[3] For the estimates that correspond to Figure 4.1 see Table A4.2.1 in Appendix 4.2.
[4] This may be the result of several factors, such as differences in constraints associated with employment decisions (e.g., no spatial constraints for males), differences in preferences regarding job attributes and labour market discrimination.

4.3 Comparison with previous estimates

The validity of the over-education estimates is considered by comparing them with estimates from previous studies examining Australian labour markets. Such comparisons, of course, cannot definitively validate the estimates because, as discussed in Section 2.3, differences in the data, samples and methods used to identify over-education are all likely to lead to differences in the estimated incidence of over-education.[5] Instead, these comparisons are regarded as merely a preliminary means for considering whether the estimates derived in this study appear reasonable (and, recall, a more reliable empirical test of their validity is then performed in Chapter 5). Table 4.3 presents the details and estimates of previous studies (along with estimates derived in this study).

Table 4.3: Previous estimates of over-education in Australia—Employed individuals (%)

Study	Data	Year	Method	Males			Females		
				Over.	*Under.*	*W-M*	*Over.*	*Under.*	*W-M*
Kler (2005)	Census [a]	1996	JA	8.0	48.0	44.0	10.0	53.0	37.0
			RM	18.0	14.0	68.0	16.0	18.0	66.0
Linsley (2005)	NLC [b]	1997	WA	28.7	17.1	54.2	25.7	21.0	53.3
Voon and Miller (2005)	Census [c]	1996	RM	15.8	13.7	70.5	13.6	18.5	67.9
Green et al. (2007)	Census [d]	1996	JA	7.4	-	-	-	-	-
	Census [d]	2001	JA	9.4	-	-	-	-	-
Kler (2007)	Census [e]	1996	JA	19.9	-	-	-	-	-
		2001	JA	21.6	-	-	-	-	-
Messinis and Olekalns (2007a)	SET [f]	1997	RM	15.5	16.2	68.3	10.9	15.1	73.9
Fleming and Kler (2008)	HILDA [g]	2001	JA	34.3	-	-	-	-	-
Mavromaras et al. (2010c)	HILDA [h]	2001	RM	19.0	-	81.0	14.0	-	86.0
		2007	RM	23.0	-	77.0	17.0	-	83.0
Jones et al. (2011)	HILDA [i]	2001-2008	RM	20.2	-	-	21.2	-	-
Mavromaras et al. (2011)	HILDA [j]	2001-2008	RM	19.0	-	66.0	22.0	-	66.0
Black (2013)	HILDA	2001	JA	16.2	10.3	73.5	21.5	8.1	70.4
		2007	JA	18.7	12.1	69.2	24.1	9.5	66.4
		2001-2008	JA	18.6	11.8	69.6	23.0	9.0	68.1

NOTES: Black (2013) refers to the estimates derived in this study (figures from Table A4.2.1 in Appendix 4.2).
[a] Sample: Australian-born individuals, aged 20–64 years. JA method based on ABS (1997) ASCO at 2-digit level.
[b] Sample: Individuals aged 18–54 years, not self-employed. 'NLC' refers to the *Negotiating the Life Course Survey*.
[c] Sample: Full-time employed individuals, aged 20–64 years.
[d] Sample: Australian-born males, aged 15–64 years. JA method based on ABS (1997) ASCO at 2-digit level.
[e] Sample: Australian-born males, with university degree, aged 15–64 years. JA method based on ABS (1997) ASCO at 2-digit level.
[f] Sample: Full-time employees. 'SET' refers to the ABS *Survey of Education and Training*.
[g] Sample: Males, aged 15 years and over. JA method based on ABS (1997) ASCO at 2-digit level.
[h] Sample: Full-time employees, with university degree, males aged 16–64 years, females aged 16–59 years. Study identified 'job mismatches' in terms of over-educated and over-skilled, but made no mention of under-education.
[i] Sample: Individuals aged 15–64 years, not self-employed.
[j] Sample: Full-time employees, males aged 16–64 years, females aged 16–59 years. Study identified 'job mismatches' in terms of over-educated and over-skilled, but made no mention of under-education.

[5] In particular, the different identification methods—the WA, JA and RM methods—tend to result in significantly different estimates of over-education, even for studies examining the same data and time period (Hartog, 2000; McGuinness, 2006). And, given the relative nature of the RM method, its estimates should not be compared to those derived using the WA and JA methods (Hartog, 2000).

Due to sample restrictions—focusing on the Australian-born, full-time employed or university graduates—the estimates of Kler (2005), Voon and Miller (2005), Green et al. (2007), Kler (2007), Messinis and Olekalns (2007a), Mavromaras et al. (2010c) and Mavromaras et al (2011) are not directly comparable to those derived in this study.[6] That leaves three studies with (potentially) comparable estimates: Linsley (2005), Fleming and Kler (2008) and Jones et al. (2011). The estimates of Linsley (2005)—28.7 per cent of males and 25.7 per cent of females over-educated—are higher than those derived in this study, particularly for males. This is most likely because they were identified using the WA method, which, when compared to the JA method, has been found to produce significantly higher estimates (Groot and Maassen van den Brink, 2000a; McGuinness, 2006; Leuven and Oosterbeek, 2011).[7] The Fleming and Kler (2008) estimate—34.3 per cent of males over-educated—is also significantly higher. While it appears the most comparable estimate (as it was also derived using HILDA Survey data and the JA method), there are two important methodological differences. First, the ABS (1997) ASCO was used to derive estimates of required education levels (rather than the ABS (2006) ANZSCO), and second the ASCO information was used at the 2-digit (rather than 4-digit) level. This use of older information regarding required education levels and examination of occupations at a less detailed level leads to the higher estimate of over-education.[8, 9] Finally, the estimates of Jones et al. (2011)—20.2 per cent of males and 21.2 per cent of females over-educated—are similar to those derived in this study. These estimates were also derived using HILDA Survey data and a similar sample (with the only difference being the inclusion of full-time students), but were identified using the RM method.[10]

---

[6] In response, of course, the same sample restrictions could be imposed here to enable comparisons. But, given the limited reliability of such comparisons (i.e., since they cannot provide definitive evidence regarding validity), this is not done.

[7] Differences in the year and sample examined (i.e., inclusion of full-time students, but exclusion of individuals aged 15–17 and 55–64 years) are also likely to explain some of the discrepancies.

[8] Recall, in preliminary analyses, this study also identified over-education using ASCO information at the 2-digit and 4-digit levels. The results—not presented, but available from the author on request—confirm that using ASCO rather than ANZSCO and using 2-digit rather than 4-digit level information (for ASCO and ANZSCO) both lead to higher estimated incidences of over-education. With regards to using ASCO, of course, the higher estimates arise because the required education levels it defines for occupations are generally lower than those defined in ANZSCO.

[9] Differences in the sample examined (i.e., inclusion of individuals aged 65 years and over, full-time students and the self-employed) are also likely to explain some of the discrepancy.

[10] Estimates from international research can also be considered. Since the incidence of over-education is likely to vary across countries, only the patterns by gender and over time are considered here. Recall, the estimates in this study indicate over-education is more prevalent among females and increasing over time. International research supports both patterns. Over-education has frequently been found to be more prevalent among females, and is a pattern observed for countries such as the US, UK, Germany and Canada (Duncan and Hoffman, 1981; Daly, Büchel and Duncan, 2000; Vahey, 2000; Dolton and Vignoles, 2000; Groot and Maassen van den Brink, 2000a; Cohn and Ng, 2000). Meanwhile, among the few studies considering changes over time in over-education rates, the evidence indicates a rising incidence of over-education, particularly in recent years (Hartog, 1997; Green et al., 1999; Felstead, Gallie, Green and Zhou, 2007; Leuven and Oosterbeek, 2011). Few studies have examined such changes because, as discussed in Section 2.3, most studies have used the RM method and its estimates are not suitable for examining changes in over-education over time (Hartog, 2000); the examination of such dynamics is therefore a suitable avenue for future research. Examination of cross-country differences in the over-education rate (rather than merely considering their patterns, as done here) is also a suitable avenue for future research; such research may highlight differences in the level of labour market rigidity across countries.

Based on comparisons with estimates from previous studies, the over-education estimates derived in this study indeed appear reasonable: in general, the estimates are within the range of all the previous estimates presented in Table 4.3, and the differences with directly comparable estimates can be explained by differences in the identification method used.[11]

4.4 Discussion and conclusion

The aim of this chapter was to determine whether there is evidence of over-education in Australian labour markets. It presented estimates of the incidence of over-education for the 2001–2008 period. The estimates indicate that approximately 20 per cent of working-age employees are over-educated in each year, while roughly 10 per cent are under-educated and 70 per cent well-matched. Also, the incidence of over-education increased over time and is higher among females: it ranged between roughly 21 and 24 per cent for females, compared with 16 to 20 per cent for males. Comparisons with results from previous studies suggested valid estimates have been derived here. This chapter, therefore, has found evidence of over-education in Australian labour markets, though the derived estimates must be further validated—this is the aim of the empirical test in Chapter 5.

[11] The estimates derived in this study are also presented by various population sub-groups in Appendix 4.2: Table A4.2.2 presents estimates by demographic characteristics, Table A4.2.3 presents estimates by human capital levels and Table A4.2.4 presents estimates by job characteristics. The results indicate over-education is more prevalent among younger individuals (aged 15 to 24 years) and those with less labour market experience, occupation tenure and employer tenure. But, it is certainly not exclusive to such individuals as there are significant proportions of older and more experienced individuals who are over-educated. Moreover, over-education is not exclusive to the most highly educated individuals (i.e., those with a Bachelor Degree or higher qualification). The highest incidences of over-education are, instead, among individuals with an education level in the middle to upper ranges of the education spectrum: those with a Certificate IV, Advanced Diploma or Bachelor Degree as their highest education level. There are also significant proportions of individuals with only a secondary education (i.e., Year 11 or Year 12) who are over-educated. Finally, results also indicate the incidence of over-education is significantly higher among NESB migrants, individuals working in certain occupations (e.g., Sales Workers) and certain industries (e.g., Retail Trade), part-time employed individuals and those working as casual employees.

Chapter 5

Over-education and under-utilised human capital

5.1 Introduction

Evidence in Chapter 4 suggests over-education exists in Australian labour markets, with estimates indicating roughly 20 per cent of working-age employees are over-educated in each year during the 2001–2008 period. But such estimates must be validated. This chapter, therefore, aims to empirically test the validity of the method used to identify over-education. In particular, it considers the research question: *do individuals who are identified as over-educated have human capital from education that is under-utilised in their current job?*

For this research question, the focus is individuals' labour productivity achieved in the workplace, where evidence of reduced labour productivity is considered indicative of individuals having human capital that is under-utilised in the job. It is assumed wages reflect the labour productivity achieved, and hence the relationship between over-education and wages is examined. Specifically, the effect over-education has on individuals' wages is estimated. Finding over-education has a negative, statistically significant effect—being over-educated *causes* an individual's wage to be lower than if they were instead well-matched—is, therefore, assumed to be evidence the individuals identified as over-educated do indeed have human capital from education that is being under-utilised. Thus, it would validate the method used to identify over-education. An important requirement for such a conclusion is that the estimates represent causal effects, not just conditional associations. However, as discussed in Section 2.4 (and Sections 5.2 and 5.3 below), estimating such causal effects is not straightforward. The concern is that unobserved individual heterogeneity (or non-random selection into over-education) may lead to biased estimates. In response, a series of empirical estimators, some of which are designed to control for unobserved individual heterogeneity, are used to derive estimates of the over-education wage penalty. The robustness of the resultant estimates is then examined, with the aim being to establish over-education wage penalty estimates that can be interpreted as causal effects.

Parametric and semi-parametric estimators are used to derive estimates. The parametric estimators—the standard approach in the over-education literature—use regression analyses to

estimate ORU earnings functions (i.e., model [2.2] in Section 2.4). In particular, estimates are derived using the pooled OLS, fixed effects and first-differences estimators. Since conditional independence (or selection on observables) is critical for the pooled OLS estimator to identify causal parameters, the ORU earnings functions are estimated with controls for detailed sets of individual characteristics. Most importantly, numerous measures are used to capture individuals' stock of human capital—their education, labour market experiences, physical and mental health, English language proficiency, family background and personality traits. But, given the difficulty in controlling for individuals' innate abilities, these pooled OLS estimates are likely biased.[1] The fixed effects and first-differences estimators are therefore used to control for such unobserved individual heterogeneity.

The semi-parametric estimators are used because, unlike regression analyses, they do not require the linear functional form assumption (Imbens and Wooldridge, 2009). This has three key benefits. First, it is no longer necessary to assume linearly conditioning on observed characteristics accounts for the potential endogeneity (or selection) bias; instead, flexible parametric specifications are used to model any non-random selection into over-education—specifically, the probability of being over-educated is modelled and predicted probabilities (or *propensity scores*) used to match over-educated and well-matched individuals, which then results in wage penalty estimates that are less likely (compared to the pooled OLS estimates) to contain endogeneity bias (Black and Smith, 2004). Second, if the wage penalty is heterogeneous across individuals, then estimates based on regression analyses, which are essentially weighted averages, do not necessarily represent the desired parameter: the mean effect of over-education among the over-educated (Borland, Tseng and Wilkins, 2005). With semi-parametric estimators, however, weights are explicitly specified to estimate this parameter (Cobb-Clark and Crossley, 2003). Third, greater consideration is given to whether the data are sufficient for deriving causal effects estimates. In particular, semi-parametric estimators highlight the importance of identifying counterfactual outcomes (e.g., the wage of an over-educated individual if they were instead well-matched), unlike regression analyses where the linearity assumption ensures (and may lead to) the identification of such counterfactuals (Black and Smith, 2004).[2, 3]

---

[1] Despite the likelihood they are biased, pooled OLS estimates are considered because they are a reasonable starting point for the analyses and are comparable with most of the over-education wage penalty estimates derived in previous studies.

[2] The estimated propensity scores also provide a means by which the issue of whether data are sufficient to derive causal effects estimates—typically expressed as the *common support assumption*—can be assessed (Cobb-Clark and Crossley, 2003). That is, comparisons of these propensity scores can indicate whether, for each over-educated individual, the data contain a sufficiently similar well-matched individual (or individuals) (in terms of observed characteristics) that can be used to estimate their counterfactual outcome and, therefore, enable the derivation of causal effects estimates.

[3] Semi-parametric estimators are further motivated by recent evidence that suggests the use of Mincer-type earnings functions to examine the relationship between education and wages, such as the ORU earnings functions, is obsolete for modern economics research (Heckman, Lochner and Todd, 2006; Heckman, Lochner and Todd, 2008). In particular, Heckman et al. (2006) found that, based on recent US data, Mincer-type earnings functions failed to produce valid estimates of the rates of return to education; thus, despite their convenience and prominence, they are no longer an accurate means for evaluating the effect education has on individuals' wages. Meanwhile, Heckman et al. (2008) encouraged the use of semi-parametric (and non-parametric) techniques to estimate the wage effects of education.

Despite these advantages, the standard semi-parametric estimator—the cross-sectional (or propensity score) matching estimator—is still reliant on conditional independence to identify causal parameters. Hence, similar to the pooled OLS estimator, the potential for endogeneity bias is minimised by using detailed sets of individual characteristics to model selection into over-education and, therefore, match over-educated and well-matched individuals. These controls, particularly the measures capturing individuals' human capital, are similar to those used in deriving the pooled OLS estimates. A further semi-parametric estimator that is used combines the propensity score matching with regression estimation (the pooled OLS estimator)—it is referred to as the combined matching and regression estimator. Such a combined approach leads to an empirical strategy that is likely to produce more reliable estimates of causal effects (Ho, Imai, King and Stuart, 2007; Imbens and Wooldridge, 2009). Ho et al. (2007) refer to it as *doubly robust*: if either the matching or the regression model is correct, but not necessarily both, then the result is consistent estimates of causal effects.[4] But it too is reliant on conditional independence to identify causal parameters. As a result, the difference-in-differences (or longitudinal) matching estimator, which essentially applies the propensity score matching estimator to longitudinal data, is also used because it is designed to control for unobserved individual heterogeneity (i.e., allows for selection on unobservables) (Heckman, Ichimura and Todd, 1997; Smith and Todd, 2005).

This chapter proceeds as follows. Section 5.2 briefly recaps existing evidence on the effect over-education has on individuals' wages. Section 5.3 uses the potential outcomes framework (or the Rubin causal model) to define the causal effect of over-education, and then outlines the empirical estimators used to derive estimates of the over-education wage penalty. Sections 5.4 and 5.5 present and discuss the results. Section 5.6 concludes the chapter.

5.2 Existing evidence

Recall from Chapter 2, many studies have considered the relationship between over-education and individuals' wages.[5] Most studies derived pooled OLS estimates based on ORU earnings functions, where the estimated over-education wage penalty is, on average, around 15 per cent. But as previously discussed, such pooled OLS estimates are likely biased. Some studies recognised the endogeneity problem and, in response, used longitudinal data and the fixed effects estimator to estimate ORU earnings functions that explicitly account for unobserved individual heterogeneity. The resultant evidence is mixed: some studies found the wage penalty remained significant and in the

[4] Combined matching and regression estimators are also among the current best practice in the econometrics of program evaluation (Imbens and Wooldridge, 2009).

[5] The vast majority, however, did not interpret their results as an empirical test of the validity of over-education. In fact, few studies have empirically tested research questions regarding the validity of over-education; most merely assume it is a valid concept and that the identification method used is also valid.

range of 5 to 20 per cent, while others found it declined to insignificance.[6] For each of these studies, however, data limit the validity of the estimates. In particular, since fixed effects estimates are reliant on the observed variation in individuals' over-education status over time, their use of the RM method and data with short panels and little variation in over-education status makes it difficult to interpret the results as causal effects estimates.[7] Therefore, due to the endogeneity bias and data limitations associated with previous estimates, there is no clear evidence on whether over-education has a causal effect on individuals' wages.[8]

This chapter contributes in several ways to the over-education literature. First, it addresses the unresolved issues regarding the validity of over-education as a labour market phenomenon and the validity of the method used to identify over-education. Also, since the empirical test performed is based on an examination of the relationship between over-education and wages, it addresses the limitation that most over-education wage penalty estimates cannot be interpreted as causal effects. The use of detailed data and sophisticated econometric techniques means more reliable estimates are contributed to the previously unclear evidence. Specifically, this chapter improves on the existing pooled OLS and propensity score matching estimates by controlling for a more extensive array of individual characteristics, particularly measures of individuals' stock of human capital (e.g., number of post-school qualifications, tenure in current occupation, tenure with current employer, years spent unemployed, recent labour market experiences, physical and mental health, English language proficiency, family background and personality), thereby reducing the potential for biased estimates. It also improves on existing fixed effects estimates as the analyses use the preferred JA method to identify over-education and longitudinal data with more observations per individual. And it is the first study to use the combined matching and regression and difference-in-differences matching estimators, where the latter technique can produce unbiased estimates even if there is non-random selection into over-education.

---

[6] But, a common finding among these studies was that fixed effects estimates were significantly smaller than pooled OLS estimates. This was interpreted as empirical evidence confirming the endogeneity bias in pooled OLS estimates.

[7] A few studies also used the propensity score matching estimator. But, as previously discussed, its reliance on conditional independence to identify causal parameters means it is not guaranteed to overcome the endogeneity problem. Ultimately, the wage penalty estimates in these studies were either similar to the pooled OLS estimates or only slightly reduced.

[8] Alternative empirical tests of the validity of over-education have also been performed. One approach used worker satisfaction levels as a proxy for individuals' labour productivity: Tsang (1987) and Tsang et al. (1991) found over-education significantly reduced satisfaction, while Büchel (2002) found no such effect. In each study, however, the estimators did not control for the potential effects of unobserved individual heterogeneity and so the resultant estimates may be biased. Moreover, Büchel (2002) compared over-educated individuals with their less educated colleagues, rather than individuals with the same education level, and so the estimates are not equivalent to the causal effect of over-education. Another approach similarly estimated ORU earnings functions but included an identifier for whether individuals believed themselves 'over-skilled' for their job; it was argued a significant reduction in the over-education wage penalty following the inclusion of the over-skilled control was evidence that over-educated individuals had under-utilised human capital (Allen and van der Velden, 2001; Di Pietro and Urwin, 2006; Allen and de Weert, 2007; Green and McIntosh, 2007). These studies found no evidence of such a reduction. But, once again, their results are likely biased due to unobserved individual heterogeneity. Therefore, similar to the evidence on the causal effect over-education has on wages, these alternative empirical tests provide no clear evidence on whether over-educated individuals have under-utilised human capital.

5.3 Framework and empirical estimators

Causality and the estimation of causal effects are contentious issues in most scientific research, including economics (Pearl, 2000; Granger, 2007; Machamer and Wolters, 2007). In this study, causal effects are defined using the potential outcomes framework (or the Rubin causal model), where a *causal effect* is the ceteris paribus change in an outcome (Y) across two states (or events) (s and s') for a particular individual (i): $Y_i(s) - Y_i(s'),\ s \neq s'$ (Rubin, 1974; Holland, 1986; Heckman, 2008).[9] Such causal effects are then estimated using parametric and semi-parametric estimators, where *conditional independence* is critical for identifying causal parameters. Conditional independence—also often referred to as *unconfoundedness*, *selection on observables*, *absence of omitted variables*, *ignorability* and *exogeneity*—means that, conditional on the set of observed individual characteristics, there are no unobserved factors that are associated with both the potential outcomes and the probability of experiencing a particular state (or event) (Cameron and Trivedi, 2005; Ho et al., 2007; Imbens and Wooldridge, 2009). This definition and estimation of causal effects is consistent with the approach of leading researchers in applied economics (see, for example, Heckman (2008), Angrist and Pischke (2009) and Imbens and Wooldridge (2009)).[10]

5.3.1 Causal effect of over-education

For each individual i, where $i = 1,\ldots, N$, let O_i be an indicator for whether currently over-educated and Y_i a random variable measuring a particular outcome (e.g., each individual's current wage). Assume each individual has two potential outcomes: Y_{1i} if over-educated and Y_{0i} if well-matched. The causal effect of being over-educated—the over-education wage penalty—is then: $\Delta_i = Y_{1i} - Y_{0i}$. Since it may vary across individuals, the mean causal effect within various sub-groups of the population can be considered. In this study, the parameter of interest is the mean effect of being over-educated rather than well-matched among all over-educated individuals—referred to as the *average treatment effect on the treated (ATT)* in the program evaluation literature—which is given by:

$$\begin{aligned} E\left[\Delta_i \mid O_i = 1\right] &= E\left[Y_{1i} - Y_{0i} \mid O_i = 1\right] \\ &= E\left[Y_{1i} \mid O_i = 1\right] - E\left[Y_{0i} \mid O_i = 1\right]. \end{aligned} \qquad [5.1]$$

The (Y_{1i}, Y_{0i}) are potential outcomes because for each individual only one can be realised. That is, the outcome observed in data (Y_i) is:

---

[9] It is implicitly assumed that the direction of causality runs from the states (or events) to the outcome (or, alternatively, that the states arise (or events occur) temporally before the outcome is realised). For this study, it is assumed that over-education affects individuals' wages or, more specifically, that the causal effect over-education has on wages represents differences in wages that are solely due to individuals being over-educated rather than well-matched.

[10] The approach is also recommended for social sciences research in general (see, for example, Morgan and Winship (2007) and Imai, Keele, Tingley and Yamamoto (2011)).

$$Y_i = O_i Y_{1i} + (1 - O_i) Y_{0i}, \qquad [5.2]$$

such that the Y_{0i} of the over-educated and Y_{1i} of the well-matched are never observed. Hence, the causal effect of over-education Δ_i is never observed, and the parameter in [5.2] cannot be identified because $E[Y_{0i} \mid O_i = 1]$ is never observed. The need to derive estimates of such counterfactuals is the fundamental problem associated with causal inference.[11]

5.3.2 Parametric estimators – Regression estimates of ORU earnings functions

Following the bulk of the over-education literature, one way to estimate the causal effect over-education has on individuals' wages is to use regression analyses to estimate ORU earnings functions. For this study using panel data on individuals, the underlying model is defined as:

$$Y_{it} = X_{it}'\beta + \delta_o O_{it} + \alpha_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 1,...,T \qquad [5.3]$$

where Y_{it} is the (log) hourly wage of individual i at time t, X_{it} is a vector of controls for individual characteristics at t (including measures of individuals' stock of human capital), β is a vector of coefficients associated with X_{it}, O_{it} is an indicator for being over-educated at t, δ_o is the estimated parameter of interest, α_i is a time-invariant random variable representing unobserved effects for individual i, and ε_{it} are idiosyncratic errors that vary across i and t (i.e., the usual regression disturbances).[12, 13] Since α_i represents factors for each individual that are unobserved in data, the model in [5.3] cannot be estimated.

If α_i contains no factors correlated with both an explanatory variable (X_{it}, O_{it}) and wages, then it is merely another unobserved factor affecting Y_{it} that is not captured by the model and subsequently absorbed into the error term. In this case, [5.3] can be re-written to produce the following estimating equation:

$$Y_{it} = X_{it}'\beta + \delta_o O_{it} + \nu_{it}, \qquad i = 1,...,N; t = 1,...,T \qquad [5.4]$$

where ν_{it} is a composite error term with $\nu_{it} = \alpha_i + \varepsilon_{it}$ and OLS estimation of [5.4] leads to pooled OLS estimates. The pooled OLS estimate of δ_o will be an unbiased and consistent estimate of the causal effect of over-education provided the above no correlation condition—the conditional

[11] A limitation of the potential outcomes framework is that estimates do not account for general equilibrium (or interaction) effects—the stable-unit-treatment-value assumption (Cobb-Clark and Crossley, 2003; Imbens and Wooldridge, 2009). Hence, causal effects are assumed independent of the number of individuals over-educated and the availability of alternative jobs, and the estimates are best regarded as the effect of changing over-education status for one individual at the margin.

[12] Unlike the ORU earnings functions defined in Chapter 2, the model in [5.3] does not contain an identifier for being under-educated. This is because the focus here is on estimating the causal effect of over-education—where it is the comparison of over-educated and well-matched individuals that is relevant—and so the observations of under-educated individuals are omitted from the analyses. For the semi-parametric estimators, attention is similarly restricted to the comparison of over-educated and well-matched individuals.

[13] The model assumes the coefficients β and δ_o are constant over time; hence, the over-education wage penalty is assumed to be constant over the eight-year period examined. This model is equivalent to a basic linear unobserved effects model or one-way error component model (Baltagi, 2001; Cameron and Trivedi, 2005).

independence (or selection on observables) assumption—holds (Heckman and Robb, 1985; Wooldridge, 2003; Black and Smith, 2004; Cameron and Trivedi, 2005).[14] But, since the error terms in [5.4] are almost certainly correlated over time for each individual, the usual OLS variance matrix (which assumes independent and identically distributed errors) will likely over-state the precision of such estimates (Cameron and Trivedi, 2005). As a result, robust (panel-corrected) standard errors must be calculated to account for this potential serial correlation in the errors of individuals with multiple observations in the data and, therefore, ensure valid statistical inference.

If α_i contains factors correlated with both an explanatory variable and wages, then pooled OLS estimates will be biased and inconsistent (Wooldridge, 2003; Cameron and Trivedi, 2005). Recall, the main concern here is factors associated with individuals' stock of human capital as each factor is likely correlated with both an individual's likelihood of being over-educated and their wage. It is therefore necessary to control for each of these factors when using regression analyses, such as the OLS estimation of [5.4], to estimate the causal effect of over-education. Hence, in this chapter, X_{it} includes numerous measures regarding individuals' human capital—the exact list of controls used for human capital (and all other individual characteristics) is presented in Section 5.4.2. A difficulty, of course, is controlling for the human capital individuals derive from their innate abilities. If the host of human capital controls used is unable to account for such factors, the pooled OLS estimates will be biased. The likely direction of this bias can be determined using a standard formula for omitted variable bias, as defined in Dolton and Silles (2008):

$$\hat{\delta}_{o,ols} = \delta_o + \frac{Cov(\alpha_i, O_{it})}{Var(O_{it})} \qquad [5.5]$$

where δ_o is the *true* causal effect of over-education. If an over-education wage penalty exists ($\delta_o < 0$) and, consistent with evidence discussed in Section 2.4, individuals' innate abilities and over-education are negatively correlated ($Cov(\alpha_i, O_{it}) < 0$), then pooled OLS estimates will be upwardly biased (i.e., larger negative numbers than the true causal effect).

When α_i is correlated with both an explanatory variable and wages, unbiased and consistent coefficient estimates can be obtained by transforming the model to eliminate α_i. One approach is to time demean [5.3], and this leads to the fixed effects estimating equation:

$$\left(Y_{it} - \overline{Y}_i\right) = \left(X_{it} - \overline{X}_i\right)'\beta + \left(O_{it} - \overline{O}_i\right)\delta_o + \left(\varepsilon_{it} - \overline{\varepsilon}_i\right) \qquad [5.6]$$

where $\overline{Y}_i, \overline{X}_i, \overline{O}_i$ and $\overline{\varepsilon}_i$ are all time averages for each individual and OLS estimation of [5.6] leads to fixed effects estimates. The fixed effects estimate of δ_o will be an unbiased and consistent estimate of the causal effect of over-education provided the regressors are strictly exogenous conditional on the

[14] Specifically, $E\left(X_{it}'\nu_{it}\right) = 0$ or $Cov\left(\nu_{it}, X_{it}\right) = 0$ must hold, such that $E\left(X_{it}'\varepsilon_{it}\right) = 0$ and $E\left(X_{it}'\alpha_i\right) = 0$.

unobserved effect (Wooldridge, 2003).[15] And, importantly, it controls for the effects of unobserved individual heterogeneity (i.e., allows for selection on unobservables).[16] Since resultant estimates are unbiased and consistent regardless of whether α_i is actually correlated with an element of X_{it} (i.e., the conditional independence assumption is relaxed), the fixed effects estimator provides causal effects estimates under weaker assumptions (Wooldridge, 2003; Cameron and Trivedi, 2005). The cost of this robustness, however, is that time-invariant factors cannot be included in the model and that the coefficient estimates for each explanatory variable are identified only by the individuals whose values change over time. Thus, the estimate of δ_o is identified only by the individuals in the data who move between the over-educated and well-matched states. This has two main implications. First, the data on variation in over-education status must be accurate.[17] Second, for the estimate to represent the causal effect, the individuals whose over-education status changes must be representative of all over-educated individuals (i.e., there must be no selection effects in the variation in over-education status). As a result, the number of individuals observed changing over-education status (or the amount of year-to-year variation) acutely affects the validity of the fixed effects estimate. Moreover, if there are few such individuals, then the causal effect will be imprecisely estimated.

An alternative way to eliminate α_i from [5.3] is to subtract a lag of each variable, and this leads to the first-differences estimating equation:

$$\left(Y_{it} - Y_{it-1}\right) = \left(X_{it} - X_{it-1}\right)'\beta + \left(O_{it} - O_{it-1}\right)\delta_o + \left(\varepsilon_{it} - \varepsilon_{it-1}\right) \qquad [5.7]$$

where differencing means the first time period observation is lost for each individual and time-invariant factors cannot be included in the model. OLS estimation of [5.7] leads to first-differences estimates. Similar to fixed effects estimate, the first-differences estimate of δ_o will be unbiased and consistent provided the regressors are strictly exogenous conditional on the unobserved effect (Cameron and Trivedi, 2005). The fixed effects and first-differences estimates should be similar and, while there is no clear preference, it is likely first-differences estimates will be less efficient because the differencing leads to estimation with fewer observations and it may reduce the variation in the regressors (Wooldridge, 2003).

---

[15] Specifically, $E\left(\varepsilon_{it} \mid X_{it}, \alpha_i\right) = 0$, which implies $Cov\left(\varepsilon_{it}, X_i\right) = 0, t = 1, ..., T$ where $X_i \equiv \left(X_{i1}, X_{i2}, ..., X_{iT}\right)$ such that for each individual the idiosyncratic error at t is uncorrelated with each regressor in all time periods.

[16] This ability to control for unobserved individual heterogeneity means the fixed effects estimates control for the factors not considered in the identification of over-education; specifically, as discussed in Section 2.3, the potential effects of individuals' innate abilities, quality of educational institution attended, qualification vintage and qualification field of study.

[17] It is for this reason that the accuracy of the observed changes in individuals' highest education level and ANZSCO occupation were scrutinised in Section 3.4.

5.3.3 Semi-parametric estimators – Matching estimates

The intuition behind the semi-parametric matching estimators is that one way to estimate the unobserved counterfactual wage for each over-educated (or *treated*) individual is to use the wage of an individual (or individuals) very similar to them (in terms of a set of observable characteristics, denoted X_i) from the group of well-matched (or *untreated*) individuals. Matching methods estimates are most commonly derived using cross-sectional data and propensity score matching of individuals. But, there are also more recent approaches where estimates are derived using a combination of propensity score matching and regression estimation, and where the use of longitudinal data leads to an enhanced propensity score matching estimator. Each of these estimators is discussed below.[18]

Cross-sectional (or propensity score) matching estimator

Given matching of the over-educated and well-matched based on X_i, which are characteristics that affect both the likelihood of being over-educated and wages, the parameter of interest is:

$$E\left[\Delta_i \mid O_i = 1, X_i\right] = E\left[Y_{1i} \mid O_i = 1, X_i\right] - E\left[Y_{0i} \mid O_i = 1, X_i\right] \quad [5.8]$$

which matching methods estimate using:

$$E\left[\Delta_i \mid O_i = 1, X_i\right] = E\left[Y_{1i} \mid O_i = 1, X_i\right] - E\left[Y_{0i} \mid O_i = 0, X_i\right]. \quad [5.9]$$

Matching methods, therefore, produce a valid estimate of the causal effect of over-education provided $E\left[Y_{0i} \mid O_i = 1, X_i\right] = E\left[Y_{0i} \mid O_i = 0, X_i\right]$. That is, after conditioning on X_i, the wages of well-matched individuals must be the same as the wages that would have occurred for over-educated individuals had they instead been well-matched. This is essentially equivalent to the conditional independence (or selection on observables) assumption. In the econometrics literature on matching, this conditional independence assumption (or *CIA*) is often more formally defined as:

$$\left(Y_{0i} \perp O_i\right) \mid X_i \quad [5.10]$$

which states that, conditional on a set of observed characteristics, wage when well-matched is independent of being over-educated. Alternatively, conditional on observed characteristics, assignment to the over-educated and well-matched states is essentially random.[19] The CIA, therefore, requires that either the assignment to the over-educated and well-matched states is a known function of observed characteristics, or the matching of over-educated to well-matched individuals is based on such a detailed set of observed characteristics that it is reasonable to assume no differences in

[18] In economics research, matching estimators are typically used in the program evaluation literature. But, they are equally suitable for other microeconomic studies seeking causal effects estimates. Black and Smith (2004) and Lee and Coelli (2010) are examples of studies that similarly use matching methods to examine labour market effects associated with education.
[19] Technically, this version of the CIA is stronger than required to identify the parameter of interest here (i.e., the ATT); instead, the previous condition of mean independence conditional on the observed characteristics would suffice (Heckman, Ichimura, Smith and Todd, 1998; Black and Smith, 2004).

unobserved characteristics exist that will affect the outcomes (Borland et al., 2005). Since the assignment (or selection) into over-education is not explicitly known, satisfaction of the CIA consequently depends on whether a sufficiently detailed set of characteristics is used to match the over-educated and well-matched individuals.

The validity of matching estimates also depends on the *common support assumption* (CSA):

$$\Pr\left(O_i = 1 \mid X_i\right) < 1 \qquad [5.11]$$

which states that for each characteristic in X_i that satisfies the CIA, there must be some individuals who are well-matched. If there are X_i for which all individuals are over-educated, then matching cannot derive an estimate of the counterfactual wage for such over-educated individuals. A conflict, therefore, exists between the CIA and CSA: to satisfy the CIA a detailed set of characteristics is required for X_i, but matching on X_i when it is of high dimension will almost certainly lead to violation of the CSA.[20] To resolve this conflict, matching is instead performed using predicted probabilities of being over-educated, or *propensity scores*, which are denoted P(X_i) and derived from a parametric model for $\Pr\left(O_i = 1 \mid X_i\right)$. This is possible because Rosenbaum and Rubin (1983) showed that if the CIA is satisfied for a particular X_i, then it will also be satisfied for P(X_i) (i.e., $\left(Y_{0i} \perp O_i\right) \mid X_i$ implies $\left(Y_{0i} \perp O_i\right) \mid P\left(X_i\right)$) (Smith and Todd, 2005). As a result, the required version of the CIA becomes mean independence conditional on propensity scores: $E\left[Y_{0i} \mid O_i = 1, P\left(X_i\right)\right] = E\left[Y_{0i} \mid O_i = 0, P\left(X_i\right)\right]$ (Heckman et al., 1998). And satisfaction of the CSA requires that for each over-educated individual there is a well-matched individual with characteristics sufficiently similar so as to produce a similar propensity score. If the CIA and CSA are satisfied, then matching can be performed to obtain estimates of the counterfactual wages of the over-educated.

In practice there are several ways these counterfactuals can be constructed, and they lead to alternative matching estimators that may produce different estimates in finite samples.[21] Letting $P_i = \Pr\left(O_i = 1 \mid X_i\right)$, matching estimators have the following general form:

$$\hat{\delta}_M = \frac{1}{N_1} \sum_{i \in I_1 \cap C_S} \left[Y_{1i} - \sum_{j \in I_0} W(i,j) Y_{0j} \right] \qquad [5.12]$$

where I_1 is the set of over-educated individuals, I_0 the set of well-matched individuals, C_S the region where the CSA is satisfied, N_1 the number of over-educated individuals in that region and then the estimated counterfactual wage for each over-educated individual is a weighted average of the wages of well-matched individuals with weights *W(i, j)* reflecting differences in the propensities P_i and P_j

[20] This is commonly referred to as the *curse of dimensionality* (Black and Smith, 2004; Borland et al., 2005).
[21] Such differences, however, only arise in the use of finite samples; in arbitrarily large samples, all matching estimators have the same asymptotic limit as they essentially reduce to exact matching (Black and Smith, 2004).

(i.e., the degree of similarity between individuals) (Smith and Todd, 2005). The various matching estimators differ in how they construct these weights (Heckman et al., 1997).

This chapter uses two of the most common matching estimators: *nearest neighbour matching* and *kernel matching*. For nearest neighbour matching, each over-educated individual is matched to the well-matched individual who has the closest propensity score, such that in [5.12] the $W(i, j) = 1$ for the well-matched individual with P_j that minimises $\left|P_i - P_j\right|$ (and then $W(i, j) = 0$ for all other well-matched individuals). Nearest neighbour matching is frequently applied because its use of a single match for each individual means it is likely the matching estimator with the least bias (Imbens and Wooldridge, 2009). Performing matching with replacement (i.e., allowing a given well-matched individual to be matched to more than one over-educated individual), as done in this chapter, further reduces the likelihood of such bias (as, on average, replacement increases the quality of the matches achieved), though it does so at the expense of precision (as it reduces the number of well-matched individuals used to construct the counterfactuals) (Smith and Todd, 2005). To ensure the CSA is satisfied, the so-called *caliper method* is used: this means over-educated individuals are only matched to well-matched individuals who have a propensity score P_j within a 5 per cent confidence interval (or two standard deviation range) of their P_i (Smith and Todd, 2005; Borland and Tseng, 2007). This reduces the possibility of poor matches and, as a result, further reduces the potential for bias, but it also means the resultant estimates do not correspond to any over-educated individuals who are unable to be matched (and subsequently dropped from the analysis).

For kernel matching (with the caliper method again used to enforce the CSA), each over-educated individual is matched to all the well-matched individuals who have P_j within a 5 per cent confidence interval of their P_i, and their counterfactual wage is calculated as the kernel-weighted average of the wages of these well-matched individuals. Such kernel weighting is designed to give more weight to well-matched individuals with P_j closer to P_i, and less to those further away. In particular, the weights are defined as:

$$W(i, j) = \frac{G^{ij}}{\sum_{j \in I_0} G^{ij}}, \text{ with } G^{ij} = G\left[\left(X_i'\hat{\beta} - X_j'\hat{\beta}\right) / a_{5\%}\right] \qquad [5.13]$$

with G^{ij} the kernel for i^{th} over-educated individual and j^{th} well-matched individual, $X_i'\hat{\beta}$ and $X_j'\hat{\beta}$ are linear predicted scores for being over-educated and $a_{5\%}$ indicates the use of the caliper method (i.e., a 5 per cent confidence interval bandwidth).[22] Compared to the nearest neighbour matching

[22] This chapter uses a biweight kernel for G^{ij} and linear predicted scores, rather than predicted probabilities, as it allows for symmetry in the selection of well-matched individuals while using the caliper method (Borland and Tseng, 2007).

estimator, the kernel matching estimator trades reduced variance (i.e., using more information to calculate counterfactuals) for increased bias (i.e., using, on average, poorer matches).

Combined matching and regression estimator

The combined matching and regression estimator is a two-stage empirical strategy. The first stage uses propensity score matching—such as the nearest neighbour and kernel matching discussed above—to match over-educated and well-matched individuals, notionally generating N_1 pairs of matched individuals. And the second stage uses this matched sample to estimate the ORU earnings function in [5.4]—hence, pooled OLS estimates of δ_o are derived using the matched sample.[23] The resultant estimates are more robust than those obtained from regression on the full sample as they are less sensitive to changes in the model specification, and they are likely to be more precisely estimated and less susceptible to bias (Ho et al., 2007). These estimates are superior because the first stage matching ensures they are based on a sample with similar covariate distributions, meaning the regression is not being used to extrapolate to outliers (Imbens and Wooldridge, 2009). Moreover, compared to the standard matching estimators that estimate causal effects using a simple mean difference in outcomes for the N_1 pairs, this combined method is able to remove any bias that may arise from discrepancies in the covariates of the matched over-educated and well-matched individuals (Imbens and Wooldridge, 2009). The combined matching and regression estimator is, therefore, considered *doubly robust* because if either the matching or the regression model is correct, but not necessarily both, then the result is consistent estimates of causal effects (Ho et al., 2007). For these reasons, it is among the current best practice in the econometrics of program evaluation (Imbens and Wooldridge, 2009). It is, however, still reliant on conditional independence (or selection on observables) to identify causal parameters.

Difference-in-differences (or longitudinal) matching estimator

The difference-in-differences (DID) matching estimator, as defined by Heckman et al. (1997) and Heckman et al. (1998), is designed to eliminate the time-invariant differences in the wages of over-educated and well-matched individuals that cross-sectional matching estimators cannot.[24] As such, it controls for unobserved individual heterogeneity (i.e., allows for selection on unobservables) (Heckman et al., 1997; Smith and Todd, 2005). The DID matching estimator essentially extends the

[23] This is similar to the local linear matching estimator described in Heckman et al. (1997) and Smith and Todd (2005).
[24] That is, following cross-sectional matching, there may still be systematic differences in the wages of over-educated and well-matched individuals (e.g., if selection into over-education is based on unobserved characteristics like individuals' innate abilities). In such situations, the conditional independence (or selection on observables) assumption is violated and, hence, the resultant cross-sectional matching estimates are biased. The DID matching estimator is designed to overcome this problem by eliminating the time-invariant differences in the wages of over-educated and well-matched individuals.

cross-sectional (propensity score) matching estimators to a setting with longitudinal data.[25] It seeks to estimate the following:

$$D_{t,t'}\left(P(X)\right) = E\left(Y_{1t} - Y_{0t'} \mid O=1, P(X)\right) - E\left(Y_{0t} - Y_{0t'} \mid O=0, P(X)\right) \qquad [5.14]$$

where t and t' represent time periods after and before individuals are observed over-educated, and the subscript i is omitted for notational convenience.[26] $D_{t,t'}\left(P(X)\right)$ is a valid estimate of the causal effect of over-education, as defined in [5.8], provided:

$$E\left(Y_{0t} - Y_{0t'} \mid O=1, P(X)\right) = E\left(Y_{0t} - Y_{0t'} \mid O=0, P(X)\right) \qquad [5.15]$$

which means that, conditional on a set of observed characteristics (or propensity scores), the wages of well-matched individuals must have evolved over the years in the same way the wages of over-educated individuals would have evolved had they not been over-educated at t—in terms of both the observable and unobservable components of wages (Blundell and Costa Dias, 2000). The DID matching estimator, as defined by Smith and Todd (2005), is then:

$$\hat{\delta}_{DDM} = \frac{1}{N_1} \sum_{i \in I_1 \cap C_S} \left[(Y_{1ti} - Y_{0t'i}) - \sum_{j \in I_0 \cap C_S} W(i,j)(Y_{0tj} - Y_{0t'j}) \right] \qquad [5.16]$$

where all variables are as previously defined (for [5.12]) and the weights $W(i, j)$ are determined by whichever cross-sectional matching estimator is used (e.g., nearest neighbour or kernel matching).

Since the assumptions necessary to derive valid causal effects estimates are considerably weaker than those required when using regression and cross-sectional matching, the DID matching estimator can be a valuable empirical tool (Heckman et al., 1998; Blundell and Costa Dias, 2000; Cobb-Clark and Crossley, 2003). Essentially, for the identification of causal effects estimates, the estimator is less demanding of data as it does not require a set X_i that eliminates selection bias, but more demanding in that it requires longitudinal data (Heckman et al., 1997). Empirical evidence on the superiority of the DID matching estimator, compared to the cross-sectional matching estimators, can be found in Heckman et al. (1997) and Smith and Todd (2005).

5.4 Empirical results

In this chapter, under-educated individuals are omitted from the analyses because estimates of the over-education wage penalty, as defined in Section 5.3, are derived from comparisons of over-educated and well-matched individuals. Analyses are also performed separately for males and females.

[25] Hence, the DID matching estimator is often referred to as the longitudinal matching estimator; it may also be called the conditional difference-in-differences estimator or the matched difference-in-differences estimator (Heckman et al., 1997; Blundell and Costa Dias, 2000; Cobb-Clark and Crossley, 2003; Doiron, 2004).

[26] Since matching is performed using propensity scores, this is technically a difference-in-differences propensity score matching estimator. See Smith and Todd (2005) and Girma and Görg (2007) for applications using such an approach.

This is because previous research has established that the wage effects of individuals' characteristics (e.g., parameter estimates in earnings functions), including the returns to education, can vary significantly by gender (Blinder, 1973; Oaxaca, 1973; Altonji and Blank, 1999). Hence, the over-education wage penalty may differ for males and females.[27] Also, given the use of longitudinal data, there are two further issues that warrant mentioning. First, due to sample attrition and individuals being not employed (or under-educated) in particular years, the analyses are performed using unbalanced panels.[28] To consider the robustness of the results, however, the analyses are replicated using balanced panels. Second, since the data span the 2001–2008 period, wages are CPI-adjusted to remove the effects of inflation and thereby enable comparisons of individuals across years. All wages are subsequently expressed in 2001-dollars (see Table A5.1.1 in Appendix 5.1 for further details on the derivation of the real wages measures).

5.4.1 Descriptive statistics

Prior to examining the parametric and semi-parametric estimates of the over-education wage penalty, descriptive statistics on individuals' wages and over-education status are considered. Table 5.1 presents means and standard deviations for the weekly and hourly wages of over-educated and well-matched individuals, along with the differences in their mean wages—elementary estimates of the over-education wage penalty.

Table 5.1: Average wages by over-education status and gender

	Males				Females			
	Over-educ.	Well-matched	*Difference*		Over-educ.	Well-matched	*Difference*	
			($)	(%)			($)	(%)
Real weekly wage ($)	710.22	974.08	-263.86	-27.1	465.99	628.27	-162.28	-25.8
	(491.25)	(596.93)			(302.35)	(379.54)		
N	4,302	17,315			5,362	16,680		
Real hourly wage ($)	18.41	22.52	-4.11	-18.3	16.47	19.60	-3.12	-15.9
	(16.05)	(15.57)			(12.16)	(13.28)		
N	4,293	17,294			5,355	16,659		

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Standard deviations reported in parentheses.
All differences are statistically significant at the 1% level.
See Table A5.1.1 in Appendix 5.1 for details on the derivation of real weekly wages and real hourly wages.

[27] Based on the average wages of over-educated and well-matched males and females (presented below in Section 5.4.1), such a difference appears to exist. An elementary empirical test—using the sample of males and females and pooled OLS to estimate an earnings function with the over-educated indicator, a female indicator and an interaction of the two (along with a constant term)—provides further support as the results (not presented, but available from the author on request) indicate the over-education wage penalty is 2.5 percentage points smaller for females, with this difference statistically significant at the 10 per cent level. Such parameter estimates, however, are likely biased (due to endogeneity problem discussed earlier). Nevertheless, performing analyses separately by gender is the more flexible approach as it does not assume the wage penalty is equal for males and females (though it may still be found to be equal). The cost of such flexibility is, of course, reduced efficiency, whereby estimations with fewer observations will likely result in less precisely estimated parameter estimates.
[28] Recall, Section 3.4.3 established that such sample attrition is unlikely to be a source of bias in results derived in this study.

The figures in Table 5.1 indicate over-educated individuals earn less than the well-matched, both before and after hours worked are taken into consideration. Over-educated males earn a weekly wage that is, on average, 27 per cent less than well-matched males, while the difference is around 26 per cent for females. Accounting for hours worked reduces the size of these differences: on average, over-educated males earn an hourly wage 18 per cent lower than well-matched males, while over-educated females earn 16 per cent less. These differences are all statistically significant at the 1 per cent level. Preliminary evidence, therefore, indicates a sizeable over-education wage penalty, which is somewhat larger for males.

Since the fixed effects, first-differences and DID matching estimates are all identified by the individuals in the data who move between the over-educated and well-matched states, the amount of variation over time in individuals' over-education status is another important consideration. Table 5.2 presents descriptive statistics on such variation. Panel A contains standard deviations, with the *between* and *within* variation distinguished—the *between* is variation across individuals, while the *within* is variation over time for each individual (Cameron and Trivedi, 2009). Standard deviations in individuals' wages are also presented, with (log) real hourly wages used as they are the wage measure most often examined in this chapter. Panel B further examines the within variation by presenting individuals' rates of transition between the over-educated and well-matched states.[29]

Table 5.2: Variation in wages and over-education status by gender

Males			Females		
A. Standard deviations					
ln(real hourly wage)	Overall	0.500	ln(real hourly wage)	Overall	0.459
(mean = 2.957)	Between	0.480	(mean = 2.833)	Between	0.440
(N = 21,459)	Within	0.238	(N = 21,850)	Within	0.256
Over-educated	Overall	0.399	Over-educated	Overall	0.429
(mean = 0.199)	Between	0.383	(mean = 0.243)	Between	0.411
(N = 21,617)	Within	0.180	(N = 22,042)	Within	0.190
B. Transitions between states					
	Over-educated $_t$	Well-matched $_t$		Over-educated $_t$	Well-matched $_t$
Over-educated $_{t-1}$	2,603	469	Over-educated $_{t-1}$	3,272	545
	(84.7%)	(15.3%)		(85.7%)	(14.3%)
Well-matched $_{t-1}$	447	12,581	Well-matched $_{t-1}$	493	12, 049
	(3.4%)	(96.6%)		(3.9%)	(96.1%)
N	16,100		N	16,359	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTE: Panel B reports row percentages in parentheses.
Number of observations is significantly lower in panel B for two reasons: (i) individuals with only one observation in the HILDA Survey data are omitted from the analyses (1,428 males and 1,479 females omitted); and, (ii) one time period (or observation) is lost for each individual in calculating the transitions.

[29] For the figures in panel B, observations may be more than one year apart for a given individual (if they attrit from the sample, are not employed or under-educated in a year between the two points). These figures are presented because the analyses are performed using unbalanced panels, and so such transitions may be used in identifying the estimates.

Results in Table 5.2 indicate far more between variation than within variation in over-education status, and more variation in wages than over-education status. Also, the over-educated and well-matched states are both highly persistent; overall only around 6 per cent of males and females are observed moving between the over-educated and well-matched states. Given this limited within variation in over-education status, a relevant concern is whether the individuals whose over-education status changes are representative of all over-educated individuals. Recall from Section 5.3, the ability to interpret the fixed effects, first-differences and DID matching estimates as causal effects is directly affected by this representativeness. Results in panel B indicate roughly 25 per cent of the males and females over-educated (in at least one year) experience a change in over-education status over the period examined.[30] But further analyses find that, in terms of observable characteristics, these sizeable sub-samples are not representative of all over-educated individuals.[31]

Compared to all over-educated individuals, the sample of individuals who change over-education status contains: a higher proportion of younger individuals (aged 15 to 29 years); a lower proportion of NESB migrants; more single individuals; less individuals with children; more highly educated individuals (i.e., those with a Bachelor Degree or higher) and less individuals with a Certificate IV, Diploma or Advanced Diploma as their highest education level; individuals with less labour market experience, occupation tenure and employer tenure (roughly three years less experience and one year less tenure); more individuals with parents who had a post-school qualification; and, individuals with more extroverted personalities. Each of these differences is statistically significant at the 5 per cent level, and typically more pronounced for males. The differences, however, are predominantly driven by the individuals who exit over-education. Specifically, analysis of the representativeness of individuals who exit over-education and the representativeness of individuals who enter over-education finds that, compared to all over-educated individuals, the sample exiting has each of the above differences, while the sample entering only has the following, relatively small differences: a lower proportion of NESB migrants; less individuals with

[30] That is, among the 3,519 males (4,310 females) over-educated at either *t* or *t–1*, there are 916 (1,038) who are observed either entering or exiting over-education.

[31] This representativeness is empirically tested using t-tests and an extensive set of individual characteristics; in particular, for each characteristic, t-tests are conducted to determine whether there are statistically significant differences in the means of the two groups (i.e., individuals whose over-education status changes vs. all over-educated individuals). The following characteristics are examined: *age*; *ethnicity*; *marital status*; *number of children*; *number of resident and non-resident children by age*; *highest education level*; *number of post-school qualifications*; *labour market experience*; *occupation tenure*; *employer tenure*; *years unemployed*; *English proficiency*; *general health rating*; *mental health rating*; *long-term health condition*; *number of jobs last year*; *proportion time employed last year*; *proportion time unemployed last year*; *State/Territory of residence*; *remoteness of area of residence*; *rest of household income*; *number of homes in past 10 years*; *father/mother had post-school qualification*; *father's/mother's ethnicity*; *father's/mother's employment status when individual 14*; *father unemployed for 6 months or more*; *number of siblings*; *whether eldest child*; *whether income support recipient*; *whether in multiple jobs*; *life satisfaction*; and, *personality traits* (i.e., *openness to experience, conscientiousness, extroversion, agreeableness, emotional stability*). The tests are conducted by gender, and additional tests then examine the representativeness of the individuals who exit over-education and the individuals who enter over-education. The results are not presented, but are available from the author on request.

Certificate IV, Diploma and Advanced Diploma highest education levels; more individuals with parents who had a post-school qualification; and, individuals with more extroverted personalities.

This evidence of selection effects in the variation in over-education status must be taken into account when examining the fixed effects, first-differences and DID matching estimates of the over-education wage penalty. In particular, the non-representativeness of the sample who change over-education status means the fixed effects and first-differences estimates are unlikely to represent causal effects. And, as discussed later, it further motivates the use of the DID matching estimator.

5.4.2 Regression estimates

Recall, estimation of model [5.4] produces the pooled OLS estimates of the over-education wage penalty. To minimise the potential for unobserved individual heterogeneity to result in biased estimates, the model controls for a detailed set of individual characteristics, with particular attention given to capturing each individual's stock of human capital. The following measures are used to directly capture individuals' human capital: *highest education level*; *number of post-school qualifications*; *labour market experience*; *tenure in current occupation*; *tenure with current employer*; *years unemployed*; *English proficiency*; *existence of a long-term health condition*; *general health rating*; and, *mental health rating*.[32] In addition, there are measures to account for factors that may indirectly affect individuals' human capital. These capture the following: recent labour market experiences (*number of jobs last year*, *proportion time employed last year*, *proportion time unemployed last year*); family background (*whether father/mother had a post-school qualification*, *ethnicity of father/mother*, *employment status of father/mother when individual aged 14*, *whether father unemployed for six months or more when individual grew up*, *number of siblings*, *whether eldest child*); and, personality traits (*openness to experience*, *conscientiousness*, *extroversion*, *agreeableness*, *emotional stability*). Standard controls for demographic characteristics are also included in the model, while measures for region of residence, rest of household income and relative mobility of individuals (using *number of homes lived in during past ten years* as a proxy) are included because each may affect the employment opportunities available to individuals and, as a result, their likelihood of being over-educated.[33] Finally, the model does not contain measures that may be affected by an individual's (current) over-education status (e.g., job characteristics, number of jobs, whether searching for a new job and life satisfaction) because such (potentially endogenous) variables would likely result in biased estimates (Ho et al., 2007).[34]

---

[32] Squared-terms for experience, occupation tenure and employer tenure (and a cubed-term for experience) capture the typically non-linear effects of experience and tenure.

[33] The complete set of controls used is listed in the Notes to Table 5.3. For further details on the explanatory variables used in this analysis see Appendix 5.1: Table A5.1.1 contains descriptions and details on variable derivations; and, Table A5.1.2 contains descriptive statistics (means and standard deviations) by over-education status and gender for the sample analysed.

[34] Previous over-education research has not always imposed such a restriction, as many studies included job characteristics in ORU earnings functions. This is a further potential source of bias in existing over-education wage penalty estimates.

As Table A5.1.2 in Appendix 5.1 indicates, some of the human capital measures—the general health rating, mental health rating, whether father/mother had a post-school qualification and the five personality measures—contain missing information for large numbers of individuals. Hence, the inclusion of such measures results in significantly reduced sample sizes for estimation. Various specifications of the model are therefore estimated: the base specification omits these variables; (three) alternative specifications add each of them separately; and, the final specification adds them all to the model (it is referred to as the *full specification*). Estimating these five specifications provides a means for examining the robustness of the results. Further sensitivity analyses are then conducted by using balanced panel samples to estimate the base and full specifications.[35] Table 5.3 presents the results—the pooled OLS estimates of the over-education wage penalty. In particular, estimates from the base and full specifications using unbalanced panels are in columns (I) and (II), while the estimates from these specifications using balanced panels are in columns (III) and (IV). Table A5.2.1 in Appendix 5.2 presents the complete set of results for the models in (I) and (II).[36]

Table 5.3: Over-education wage penalty—ORU earnings functions, pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
Over-educated	-0.151**	-0.151**	-0.165**	-0.122**	-0.163**	-0.167**	-0.171**	-0.196**
	(0.012)	(0.015)	(0.025)	(0.026)	(0.010)	(0.012)	(0.025)	(0.031)
Health ratings, parent education and personality measures	No	Yes	No	Yes	No	Yes	No	Yes
Balanced panel sample	No	No	Yes	Yes	No	No	Yes	Yes
N	20,288	15,146	6,688	5,096	20,806	16,357	6,200	4,672
(No. individuals)	(5,231)	(3,260)	(836)	(637)	(5,370)	(3,637)	(775)	(584)
R-squared	0.3051	0.3065	0.3083	0.3222	0.2924	0.2960	0.3319	0.3414

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.
All models contain the following controls: *human capital measures* (highest education, multiple qualifications, experience, occupation tenure, employer tenure, years unemployed, English proficiency, long-term health condition); *recent labour market experiences* (number of jobs last year, proportion time employed last year, proportion time unemployed last year); *demographic characteristics* (age, ethnicity, marital status, number of children, number of resident and non-resident children by age, rest of household income, State/Territory of residence, remoteness of area of residence, number of homes lived in during past ten years); *family background* (father's/mother's ethnicity, father's/mother's employment status when individual aged 14, father unemployed for six months or more, number of siblings, eldest child); and, *year dummies*.
See Table A5.2.1 in Appendix 5.2 for the complete set of results for models (I) and (II).

The results are largely as expected, and coincide with standard findings in the estimation of wage equations. For instance, as approximated by R-squared measures, the models explain around

[35] In addition to being a survey respondent and employed (and not under-educated) in each year, these balanced panels require each individual to have non-missing values for all variables used in the model.
[36] Complete results for models (III) and (IV) are not presented, but are available from the author on request.

30 per cent of the variation in individuals' wages; this explanatory power may seem relatively low, but is typical in empirical studies of wages (Bowles, Gintis and Osborne, 2001). Most human capital measures are statistically significant and exert the usual effects: higher education levels, more experience and occupation tenure all increase wages, while more time unemployed, poor English proficiency and long-term health conditions reduce wages. The effects of personality measures, however, are somewhat more novel. For all individuals, greater conscientiousness is found to increase wages, while being a more agreeable person leads to lower wages. And, for females, a more extroverted personality is found to result in higher wages.

With regards to the effect of over-education, the pooled OLS estimates all indicate that being over-educated reduces individuals' wages by a statistically significant and sizeable amount. This wage penalty is roughly 15 per cent for males and 16 to 17 per cent for females. The estimates are robust across the specifications using unbalanced panels, and remain significant in the specifications using balanced panels, though the size of the effects varies somewhat. In addition, they are similar to the pooled OLS estimates in the over-education literature.[37] But, as previously discussed, it is likely these estimates are upwardly biased because the models, despite containing numerous measures for human capital, are unable to account for the human capital individuals derive from their innate abilities. The models in [5.6] and [5.7] are therefore estimated in attempts to control for such unobserved individual heterogeneity and, thereby, produce the fixed effects and first-differences estimates of the over-education wage penalty. Table 5.4 presents these estimates. The models in (I) have the same specification as the base model used to derive the pooled OLS estimates, except of course the time-invariant measures are omitted (i.e., ethnicity and family background measures). Since the fixed effects and first-differences estimators rely on within variation to identify parameter estimates, the specification in (II) seeks to improve the models by removing variables with little within variation.[38] These two models are also estimated using balanced panels, and their results presented in columns (III) and (IV).[39]

[37] Pooled OLS estimates were also derived for a model similar to [2.1]—specifically, an ORU earnings function that controls for required education level rather than highest education level. Consistent with previous results in the literature, the estimates indicate that over-educated individuals earn more than their co-workers who have the (lower) required education for the job (i.e., individuals who are well-matched and employed the same job). In particular, this wage premium is, on average, approximately 10 per cent for over-educated males and 7 per cent for over-educated females. The results from these additional ORU earnings functions are not presented, but are available from the author on request.

[38] Similar to the analysis in panel A of Table 5.2, standard deviations were examined in order to determine which variables have little within variation. For the list of variables omitted see the Notes to Table 5.4.

[39] Table A5.2.2 in Appendix 5.2 presents the complete set of results for model (I) using fixed effects and first-differences estimators. Complete results for models (II)–(IV) are not presented, but are available from the author on request.

Table 5.4: Over-education wage penalty—ORU earnings functions, fixed effects and first-differences estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
A. Fixed effects								
Over-educated	-0.036*	-0.039**	-0.071**	-0.070**	-0.060**	-0.060**	-0.041*	-0.038
	(0.015)	(0.014)	(0.024)	(0.024)	(0.014)	(0.013)	(0.021)	(0.020)
Omit variables with little within variation	No	Yes	No	Yes	No	Yes	No	Yes
Balanced panel sample	No	No	Yes	Yes	No	No	Yes	Yes
N	20,636	20,664	6,768	6,824	21,035	21,069	6,240	6,264
(No. individuals)	(5,323)	(5,327)	(846)	(853)	(5,449)	(5,459)	(780)	(783)
Within R-squared	0.0674	0.0557	0.0930	0.0768	0.0421	0.0357	0.0553	0.0458
Between R-squared	0.2060	0.1885	0.0777	0.0709	0.1591	0.0779	0.0004	0.0129
Overall R-squared	0.1616	0.1405	0.0699	0.0669	0.1136	0.0594	0.0007	0.0169
B. First-differences								
Over-educated	-0.043*	-0.043*	-0.058*	-0.061*	-0.065**	-0.062**	-0.061*	-0.055*
	(0.020)	(0.019)	(0.028)	(0.027)	(0.018)	(0.017)	(0.024)	(0.024)
Omit variables with little within variation	No	Yes	No	Yes	No	Yes	No	Yes
Balanced panel sample	No	No	Yes	Yes	No	No	Yes	Yes
N	13,959	13,986	5,922	5,971	13,981	14,006	5,460	5,481
(No. individuals)	(3,659)	(3,662)	(846)	(853)	(3,765)	(3,769)	(780)	(783)
R-squared	0.0154	0.0100	0.0215	0.0165	0.0130	0.0091	0.0181	0.0113

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.
Models (I) and (III) contain the following categories of controls: *human capital measures; recent labour market experiences*; *demographic characteristics* (excluding ethnicity); and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories. The variables with little within variation omitted from models (II) and (IV) are: highest education; experience; occupation tenure; employer tenure; English proficiency; State/Territory of residence; remoteness of area of residence; and, number of homes lived in during past 10 years.
See Table A5.2.2 in Appendix 5.2 for the complete set of results for model (I).

As expected, the fixed effects and first-differences estimates are similar, and the fixed effects estimates are more precisely estimated. Both also indicate that over-education has a negative, statistically significant effect on individuals' wages. Specifically, based on the unbalanced panels, the over-education wage penalty is roughly 4 per cent for males and 6 per cent for females. These effects are significantly smaller in magnitude than the pooled OLS estimates, which can be interpreted as empirical evidence that the pooled OLS estimates are indeed upwardly biased. A concern, however, is that given the previous evidence of selection effects in the variation in over-education status—since the individuals who identify these wage penalty estimates are not representative of all over-educated individuals—the fixed effects and first-differences estimates are unlikely to represent causal effects. In particular, these estimates may be downwardly biased because the sub-samples that identify them contain higher proportions of younger and more highly educated individuals, and such individuals,

given the empirical evidence of significantly less wage dispersion among them (Katz and Autor, 1999; Neal and Rosen, 2000), are likely to experience a smaller wage penalty (compared to the average among all over-educated individuals). That is, since there is less variation in their wages, the differences between wages when well-matched and wages when over-educated—the over-education wage penalty—are likely smaller. Also, since the selection effects are more pronounced among males, there is likely more downward bias in the fixed effects and first-differences estimates for males.[40]

5.4.3 Cross-sectional matching estimates

When exact matching on all characteristics in X_i is not possible, exact matching on the factors considered critical determinants of the outcome may be vital for the derivation of valid causal effects estimates (Card and Sullivan, 1988; Rubin and Thomas, 2000; Borland and Tseng, 2007). Given the importance of education in wage determination and the identification of over-education, exact matching on each individual's highest education level is considered necessary. And, consistent with the argument for estimating separate earnings functions for males and females, exact matching is also performed on gender. Therefore, based on estimated propensity scores, over-educated males are matched to well-matched males with the same highest education level, and similarly so for females. This is referred to as *quasi-exact matching* (Borland and Tseng, 2007). In this chapter, logit models for the probability of being over-educated, $\Pr(O_i = 1 | X_i)$, are estimated to derive the propensity scores, and separate models are estimated for males and females. Since these propensity scores are used to match over-educated and well-matched individuals, the exact specifications of the logit models is critically important.

To determine the appropriate specifications for these models, an algorithm similar to that defined by Dehejia and Wahba (2002) is used. It entails first estimating preliminary models with controls for a detailed set of individual characteristics that are likely to affect both the probability of being over-educated and individuals' wages. The controls are essentially the same as those used in the base specification of the ORU earnings functions that produce the pooled OLS estimates, with measures for human capital, recent labour market experiences, demographic characteristics and family background.[41] Recall, such a detailed set of controls is used in attempts to satisfy the CIA.

---

[40] The fixed effects and first-differences estimates are robust across specifications using unbalanced panels, but differences arise when balanced panels are examined: estimates increase for males and decrease for females. The differences are most pronounced when using the fixed effects estimator, with the wage penalty estimates roughly twice as large for males and one-third smaller for females. These differences arise because the balanced panels contain far fewer individuals who change over-education status. Despite this lack of robustness regarding the exact size of the wage penalty, the general result that over-education significantly reduces individuals' wages remains.

[41] Extensive controls for labour market history (*experience*, *occupation tenure*, *employer tenure*, *years unemployed*, *number of jobs last year* and *proportion time employed/unemployed last year*) and local labour market conditions (*State/Territory of residence*, *remoteness of area of residence* and *number of homes lived in during past ten years* (to proxy willingness to move for employment)) are included as Heckman et al. (1998) argue that controlling for such factors is likely necessary to ensure estimates represent causal effects.

Propensity scores are then derived from the models and used to perform a series of so-called *balancing tests*, which seek to determine whether conditioning on these propensity scores leads to balance—no statistically significant differences—in the (mean) characteristics of the groups of over-educated and well-matched individuals. This serves as an empirical test of whether the CIA appears satisfied. Given the importance of achieving such balance, it is necessary to perform several different balancing tests (Ho et al., 2007). This chapter uses three.[42] Where imbalance is found, the specifications are modified to include interaction and polynomial terms for characteristics, particularly those that failed the balancing tests. The logit models are then re-estimated, propensity scores re-calculated and balancing tests repeated. This process continues until specifications are found that produce acceptable balancing tests results.[43] Table A5.3.1 in Appendix 5.3 contains the logit model results with the chosen specifications, and Table A5.3.2 presents the corresponding balancing tests results.[44]

Given the estimated propensity scores, it is also important to consider whether the CSA is satisfied. That is, whether sufficient overlap exists in the propensity scores of over-educated and well-matched individuals to enable suitable comparisons and, thereby, produce valid causal effects estimates. This is done by comparing the distributions of propensity scores for the two groups. Figure 5.1 presents such comparisons.[45] The graphs clearly indicate that well-matched individuals are

---

[42] The first is a balancing test defined by Dehejia and Wahba (2002): the propensity scores are used to sort individuals into 40 strata (i.e., 0.0-0.025, 0.025-0.05, …, 0.975-1.0) and then, for each characteristic in the specification, t-tests are conducted to determine whether there are statistically significant differences in the means of the over-educated and well-matched groups, and then a Hotelling T-test is conducted to determine whether these differences are jointly statistically significant. The second is a balancing test suggested by Smith and Todd (2005): the propensity scores are used to perform the (nearest neighbour and kernel) matching and then, for each characteristic in the specification, t-tests are conducted to determine whether there are statistically significant differences in the means of the over-educated and well-matched groups in the post-matching samples, and then a Hotelling T-test is conducted on the joint statistical significance of these differences. The third is a balancing test from Ho et al. (2007): this is identical to the Smith and Todd (2005) test, except differences in the means of the over-educated and well-matched groups are now deemed significant (or large enough to warrant adjusting the specification) if they exceed one-quarter of the standard deviation of the variable in the pre-matching samples.

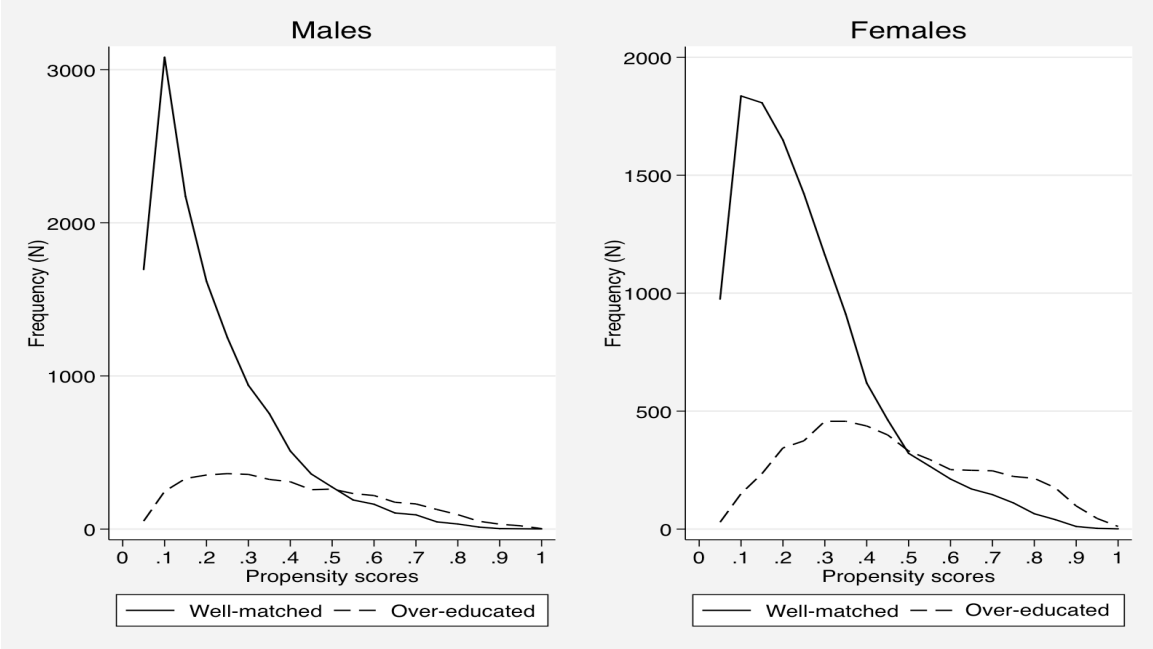
[43] The final specifications were chosen as they: (i) result in no characteristics with significant differences in Ho et al. (2007) balancing tests; (ii) result in no jointly significant differences in Smith and Todd (2005) balancing tests; (iii) minimise the number of characteristics with significant differences in Smith and Todd (2005) balancing tests; and, (iv) minimise the number of strata with jointly significant differences in Dehejia and Wahba (2002) balancing tests—the strata with jointly significant differences are those for propensity scores at the two extremes: low propensity scores because there are few over-educated in the strata and high propensity scores because there are few well-matched in the strata. The final specifications include various interaction terms, based on interacting the following: *age* and *number of children*; *age* and *experience*; *age* and *divorced*; *age* and *English proficiency*; *ethnicity* and *number of homes lived in during past ten years*; *years unemployed* and *age*, *State/Territory of residence*, *remoteness of area of residence*, *number of homes lived in during past ten years*; and, *proportion time employed last year* and *age*, *State/Territory of residence*.

[44] As indicated by Heckman et al. (1997) and Smith and Todd (2005), within-sample correct prediction rates can also be used to assess the quality of the estimated propensity scores. Specifically, the propensity scores are used to predict whether an individual is over-educated, where the actual proportions of males and females over-educated are used as the cut-off values (e.g., a male is predicted to be over-educated if their propensity score is greater than (or equal to) the proportion of males in the sample over-educated). For males, 70.2 per cent of the well-matched are correctly predicted to be well-matched and 69.8 per cent of the over-educated are correctly predicted to be over-educated. Meanwhile, the correct prediction rates for females are 71.3 per cent for being well-matched and 70.1 per cent for being over-educated. The logit models and the resultant propensity scores, therefore, are reasonably accurate in predicting the probability of being over-educated.

[45] Figure 5.1 presents the distributions of propensity scores with frequency counts rather than densities on the vertical axis because, for this analysis, results using densities can be misleading. That is, since this chapter seeks ATT estimates, the data

concentrated at much lower propensity scores than the over-educated. Nevertheless, across the distributions of propensity scores there appears to be sufficient numbers of well-matched individuals to perform both the nearest neighbour and kernel matching, though the one exception may be for over-educated females with propensity scores that exceed 0.7.[46]

Figure 5.1: Distribution of propensity scores by over-education status and gender



SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

Based on the estimated propensity scores, the quasi-exact matching using nearest neighbour and kernel matching is performed. The quality of such matching is assessed from two perspectives: (i) the number of over-educated individuals successfully matched; and, (ii) the number of well-matched individuals used in the matching, with this reliance considered across the propensity score distributions. Almost all over-educated individuals are successfully matched: only 0.6 per cent of over-educated males and 0.4 per cent of over-educated females cannot be matched. This means the representativeness of the matching estimates is unlikely to be adversely affected (i.e., the estimates almost certainly represent the mean effect of over-education among the over-educated (or the ATT)).

need only contain a sufficient number of comparable well-matched individuals to match to the over-educated. Hence, the frequency counts are the relevant factor. Also, since the data contain far more well-matched than over-educated individuals, examinations in terms of densities would likely indicate insufficient overlap when, in actuality, the small fractions of well-matched individuals at certain propensity scores may provide sufficient numbers to perform the matching. The results in Figure 5.1 and Figure A5.3.1 (in Appendix 5.3) confirm that this is indeed the case.

[46] This is, however, merely an elementary examination of the overlap issue. Since quasi-exact matching is performed on highest education level and gender, it is the distributions of propensity scores within each highest education-gender group that are relevant. But, as there are 20 such groups (see Table A5.3.3 in Appendix 5.3), the graphs in Figure 5.1 are presented to provide an overview. Ultimately, the overlap issue is dealt with during the quasi-exact matching (i.e., if necessary, data are trimmed to enforce the CSA), where the relevant distributions of propensity scores are taken into consideration.

Also, as Table A5.3.3 in Appendix 5.3 indicates, sizeable numbers of well-matched individuals are used in the matching, particularly, as expected, for the kernel matching.[47] Further analysis, presented in Table A5.3.4 in Appendix 5.3, shows that the reliance on well-matched individuals increases for higher propensity scores, and in a few categories it is potentially too high. But, these categories contain only small percentages of all over-educated individuals.[48] The matching estimates, therefore, do not appear to be overly reliant on small numbers of well-matched individuals to identify the counterfactual wages of over-educated individuals. Table 5.5 presents the matching estimates of the over-education wage penalty. In particular, panel A contains the nearest neighbour matching estimates and panel B the kernel matching estimates.

Table 5.5: Over-education wage penalty—Cross-sectional matching estimates

	Males				Females			
	Over-educ.	Well-matched	*Difference*		Over-educ.	Well-matched	*Difference*	
			($)	(%)			($)	(%)
A. Nearest neighbour matching								
Real hourly wage	18.47	21.99	-3.52**	-0.160**	16.59	20.06	-3.46**	-0.173**
			(0.418)	(0.016)			(0.402)	(0.017)
N			6,624				8,030	
B. Kernel matching								
Real hourly wage	18.51	21.67	-3.16**	-0.146**	16.61	19.50	-2.89**	-0.148**
			(0.211)	(0.009)			(0.238)	(0.011)
N			17,052				16,915	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Bootstrap standard errors based on 1,000 replications reported in parentheses.

The matching estimates all indicate that over-education has a negative, statistically significant effect on individuals' wages. This wage penalty is estimated at around 15 to 16 per cent for males and 15 to 17 per cent for females. The kernel matching estimates are somewhat lower than the nearest neighbour matching estimates and, as expected, more precisely estimated.[49] These matching estimates are essentially identical to the pooled OLS estimates. This means the relaxation of the linear

[47] Table A5.3.3 presents the number of over-educated and well-matched individuals in the highest education-gender groups pre- and post-matching (for both the nearest neighbour and kernel matching). The post-matching samples indicate the number of well-matched used in the matching for each group. Differences in the numbers of over-educated individuals in the pre- and post-matching samples, however, do not represent the number of over-educated unable to be successfully matched because the table does not distinguish between those omitted due to undefined propensity scores (i.e., individuals with missing information for an explanatory variable in the logit models) and those omitted to enforce the CSA (i.e., the individuals who cannot be matched as the data do not contain a sufficiently similar well-matched individual).
[48] Table A5.3.4 considers the issue of reliance on well-matched individuals more closely: based on the nearest neighbour matching, it presents the number of over-educated and well-matched individuals in the post-matching samples by deciles of the propensity score distributions, along with the average number of times the well-matched individuals in each decile are used in the matching. The reliance is potentially too high for males with propensity scores between 0.8 and 0.9 and for females with propensity scores between 0.8 and 1.0. This analysis was replicated using the post-matching samples from the kernel matching: the results (not presented, but available from the author on request) are qualitatively similar.
[49] But, as stated in Section 5.3.3, increased bias is the likely cost of such increased precision.

functional form assumption associated with regression analyses—or, more specifically, using flexible parametric specifications to model non-random selection into over-education, defining weights to ensure estimates represent the mean effect among the over-educated and ensuring the CSA is satisfied—has little effect on the wage penalty estimates. And, more importantly, the matching does not successfully control for the unobserved individual heterogeneity, and so it is likely these matching estimates are also upwardly biased.

5.4.4 Combined matching and regression estimates

Recall, the combined matching and regression estimates are pooled OLS estimates derived using matched samples. Thus, the post-matching samples from the previous sub-section (for both the nearest neighbour and kernel matching) are used to estimate model [5.4]. As in Section 5.4.2, two specifications of the model are estimated: the base specification (I) and the full specification (II) (containing the additional measures for health, parent education and personality, which lead to reduced sample sizes for estimation). Table 5.6 presents the wage penalty estimates.[50]

Table 5.6: Over-education wage penalty—ORU earnings functions, combined matching and pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*		Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*	
	(I)	(II)	(I)	(II)	(I)	(II)	(I)	(II)
Over-educated	-0.149**	-0.149**	-0.136**	-0.141**	-0.176**	-0.181**	-0.160**	-0.166**
	(0.014)	(0.016)	(0.012)	(0.014)	(0.014)	(0.014)	(0.011)	(0.012)
Health ratings, parent education and personality measures	No	Yes	No	Yes	No	Yes	No	Yes
N	7,868	5,707	16,938	12,842	9,970	7,563	16,776	13,199
(No. individuals)	(2,860)	(1,909)	(4,325)	(2,765)	(3,237)	(2,274)	(4,368)	(2,979)
R-squared	0.3484	0.3585	0.3272	0.3383	0.2687	0.2910	0.2611	0.2786

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.
All models contain the following categories of controls: *human capital measures*; *recent labour market experiences*; *demographic characteristics*; *family background*; and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories.
See Table A5.4.1 in Appendix 5.4 for the complete set of results for model (I).

The combined matching and regression estimates all point to a statistically significant over-education wage penalty. In particular, it is estimated at approximately 14 to 15 per cent for males and 16 to 18 per cent for females. And, as expected, the estimates are robust across the two specifications

[50] Table A5.4.1 in Appendix 5.4 presents the complete results for model (I). Complete results for model (II) are not presented, but are available from the author on request.

of the model. These wage penalty estimates are very similar to the pooled OLS and (cross-sectional) matching estimates. This means there are no outliers in the data that significantly affect the pooled OLS estimates (i.e., regression models are not being used to extrapolate to outliers) and there are no significant discrepancies in the covariates of the matched samples of over-educated and well-matched individuals that bias the matching estimates. It also means these combined matching and regression estimates are likely upwardly biased. But this is not surprising given neither the matching nor the regression models are found to successfully control for the unobserved individual heterogeneity.

5.4.5 Difference-in-differences matching estimates

Identification of the difference-in-differences (DID) matching estimates, as defined in [5.16], requires two observations for each of the over-educated and well-matched individuals examined.[51] Two-year panels of data are therefore constructed using the adjacent waves of the HILDA Survey data (i.e., W1–W2, W2–W3, etc).[52] Following Heckman et al. (1997), matching is performed in the second time period; thus, individuals over-educated at time t are matched to individuals well-matched at t, and then the difference in their changes in wages (between t and t') are used to estimate the causal effect of being over-educated at t.[53] This matching is similar to the cross-sectional matching previously performed, whereby: quasi-exact matching is performed on individuals' highest education level and gender; logit models for the probability over-educated (at t) are estimated to derive propensity scores, and separate models are estimated for males and females; specifications of the logit models are determined using the Dehejia and Wahba (2002) algorithm that considers results from balancing tests; and, the caliper method is used during matching to enforce the CSA.

In contrast to previous applications of the DID matching estimator, the study of over-education has the important distinction that individuals are not limited to experiencing the *treatment* (i.e., being over-educated) in just one time period.[54] Hence, it is possible some of the individuals over-educated at t were also over-educated at t', while others were well-matched at t'. And the same applies to the individuals well-matched at t. This has serious implications for whether the identifying assumption in [5.15] holds. Recall, the DID matching estimator produces valid causal effects estimates provided that, conditional on observed characteristics, the wages of individuals well-

[51] Recall, this is because the outcome considered is the changes in individuals' wages between two time periods. In this chapter, the change in the log of hourly wages is examined because then, similar to the use of logs in regression analyses, the differences across time periods represent percentage changes in wages.
[52] Due to the attrition in the HILDA Survey data, there is a loss of observations and so, compared to the previous analyses, the DID matching estimates are derived using smaller sample sizes. There are also sample restrictions to ensure individuals are employed in both years and not under-educated in either year.
[53] Recall, t' represents the first time period and t the second time period.
[54] Previous studies that use the DID matching estimator have tended to examine one-off treatments or events, such as participation in social programs (e.g., job training) (e.g., Heckman et al., 1997; Smith and Todd, 2005), government policy changes (e.g., Doiron, 2004) and firms becoming owned by foreign multinational enterprises (e.g., Girma and Görg, 2007).

matched at t have evolved over the two years in the same way the wages of individuals over-educated at t would have evolved had they not been over-educated at t. However, since the over-educated and well-matched states are both highly persistent, the matching performed at t essentially amounts to matching individuals over-educated in both periods ($O_{t'}$, O_t) with individuals well-matched in both periods ($WM_{t'}$, WM_t). And, as a result, [5.15] is likely violated because it effectively reduces to assuming the wage changes of individuals well-matched in both periods ($WM_{t'}$, WM_t) are equal to those of individuals over-educated at t' and well-matched at t ($O_{t'}$, WM_t) (i.e., the wage changes of individuals well-matched in both periods adequately represent the unobserved counterfactuals for individuals over-educated at t); an assumption which seems unlikely to hold.[55]

In using the DID matching estimator, therefore, quasi-exact matching must be performed on individuals' over-education status at t'. Thus, the matching is performed separately for two groups: (i) individuals over-educated at t' (i.e., ($O_{t'}$, O_t) matched to ($O_{t'}$, WM_t)); and, (ii) individuals well-matched at t' (i.e., ($WM_{t'}$, O_t) matched to ($WM_{t'}$, WM_t)). In both cases the identifying assumption is now likely to hold. Sample sizes in the two-year panels, however, mean quasi-exact matching on highest education level, gender and over-education status at t' is infeasible.[56] Instead, quasi-exact matching is performed only on over-education status at t'. Panel B of Table 5.7 presents these DID matching estimates, while panel A contains the preliminary (and likely biased) estimates that overlook over-education status at t' (but perform quasi-exact matching on highest education and gender).

The results in panel A suggest over-education has only a relatively minor negative effect on individuals' wages. This wage penalty is roughly 1 to 2 per cent for males and 3 per cent for females, and only statistically significant for females. But, as discussed above, it is likely these estimates are (downwardly) biased. On the other hand, estimates in panel B are larger in magnitude, and significantly so for females. In particular, the wage penalty is estimated at around 2 to 4 per cent for males and 8 to 9 per cent for females, though once again it is only statistically significant for females. The estimates in panel B, however, are also likely biased because the balancing tests indicate significant imbalance—statistically significant differences in the (mean) characteristics of the matched samples of over-educated and well-matched individuals—and specifications of the logit models that

---

[55] Moreover, with regards to calculation of the DID matching estimates, matching individuals over-educated in both periods with those well-matched in both periods is likely to result in downwardly biased estimates of the over-education wage penalty (i.e., smaller negative numbers than the true causal effect). This is because the estimator, as defined in [5.16] (and represented below with the subscripts i, j and notation concerning use of matching weights omitted for convenience), effectively calculates $[(Y_{1t} - Y_{1t'}) - (Y_{0t} - Y_{0t'})]$ instead of the desired $[(Y_{1t} - Y_{0t'}) - (Y_{0t} - Y_{0t'})]$, where it is likely that $(Y_{1t} - Y_{1t'}) > (Y_{1t} - Y_{0t'})$. Hence, the likely change (decline) in wages associated with the move from well-matched at t' to over-educated at t is being under-estimated, which leads to DID matching estimates that are downwardly biased.

[56] This is because it results in 40 groups—highest education (10) x gender (2) x over-education status at t' (2)—by which to perform propensity score matching and the available samples are too small for such quasi-exact matching. Furthermore, the resultant estimates are found to be highly unreliable as all three balancing tests indicate that, for almost all characteristics, statistically significant differences exist in the means of the matched samples of over-educated and well-matched individuals. These estimates and balancing tests results are not presented, but are available from the author on request.

produce acceptable balancing tests results could not be derived.[57] Insufficient sample sizes are the likely reason for the bias. Specifically, given the persistence of over-education, for the matching of group (i) (those over-educated at *t'*) there are few individuals well-matched at *t* who can be matched to the many individuals over-educated at *t*.[58] As a result, the estimates for this group are unreliable as they are based on the matching of a large *treatment group* to a small *control group*.

Table 5.7: Over-education wage penalty—Difference-in-differences matching estimates

	Males		Females	
	DID matching (*Nearest neighbour*)	DID matching (*Kernel*)	DID matching (*Nearest neighbour*)	DID matching (*Kernel*)
A. Quasi-exact matching on highest education level and gender				
Difference in (log) real hourly wage	-0.012	-0.019	-0.029	-0.026**
	(0.019)	(0.011)	(0.016)	(0.010)
N	4,337	11,814	5,047	11,458
(No. over-educated)	(2,557)	(2,541)	(3,126)	(3,112)
B. Quasi-exact matching on over-education status at *t'*				
Difference in (log) real hourly wage	-0.018	-0.036	-0.094*	-0.078**
	(0.043)	(0.033)	(0.032)	(0.024)
N	3,098	11,887	3,716	11,659
(No. over-educated)	(2,557)	(2,552)	(3,126)	(3,116)
C. Restricted sample: Individuals well-matched at *t'*				
Difference in (log) real hourly wage	-0.093*	-0.097**	-0.071*	-0.066*
	(0.038)	(0.027)	(0.032)	(0.027)
N	516	8,471	530	7,656
(No. over-educated)	(281)	(285)	(292)	(293)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Bootstrap standard errors based on 1,000 replications reported in parentheses.
Estimates in panel C derived with quasi-exact matching on highest education level and gender.

An alternative is to omit the individuals in group (i) and derive DID matching estimates based solely on those in group (ii)—the individuals well-matched at *t'*. This is a viable option because, despite persistence again leading to groups of vastly different sizes, it involves matching a small *treatment group* (the over-educated at *t*) to a large *control group* (the well-matched at *t*) and such comparisons are sufficient for the identification of the ATT (the causal effect of over-education). Also, for the results to be valid causal effects estimates, this sample of individuals who enter over-education must be representative of all over-educated individuals. Recall from Section 5.4.1, supplementary analyses find that, in terms of observed characteristics, these individuals are approximately representative of all over-educated individuals—only relatively small differences exist,

[57] Results from the balancing tests are not presented, but are available from the author on request.
[58] In the two-year panels of HILDA Survey data, sample sizes for the quasi-exact matching on over-education status at *t'* are as follows: group (i) contains 6,005 individuals, of which 5,342 are over-educated at *t* and 663 are well-matched at *t*; and, group (ii) contains 18,520 individuals, of which 619 are over-educated at *t* and 17,901 are well-matched at *t*.

whereby this sub-sample contains: fewer NESB migrants; fewer individuals with Certificate IV, Diploma and Advanced Diploma highest education levels; more individuals with parents who had a post-school qualification; and, individuals with more extroverted personalities.[59] Panel C of Table 5.7 presents the estimates.[60] The results all indicate over-education has a negative, statistically significant effect on wages, though the estimates using kernel matching may be unreliable as balancing tests indicate significant imbalance.[61] Hence, the wage penalty is estimated at roughly 9 per cent for males and 7 per cent for females. These estimates are significantly smaller than the (upwardly biased) pooled OLS, cross-sectional matching and combined matching and regression estimates, and larger than the (downwardly biased) fixed effects and first-differences estimates (though only marginally so for females). Such results, along with the previously discussed advantages of the empirical estimator, mean these DID matching estimates are considered the most accurate estimates of the over-education wage penalty derived in this chapter (i.e., those most likely to represent causal effects).

5.5 Synthesis of results

The empirical analyses in this chapter sought to determine whether over-education has a causal effect on individuals' wages; specifically, whether an over-education wage penalty exists. Estimating such causal effects is difficult because unobserved individual heterogeneity (or non-random selection into over-education) may lead to biased estimates. In response, a series of parametric and semi-parametric estimators, some of which are designed to control for unobserved individual heterogeneity, have been used to derive over-education wage penalty estimates. And now, based on examination of these estimates, the aim is to establish which, if any, can be interpreted as causal effects. Table 5.8 presents the key estimates. In particular, the first row contains the differences in the mean hourly wages of over-educated and well-matched individuals—the elementary estimates of the over-education wage penalty. The subsequent two panels contain the more sophisticated estimates, where the distinction is based on whether the conditional independence assumption is required for the empirical estimator to

[59] Given the focus on recent entrants to over-education, it must also be assumed that the effect of over-education does not vary by individuals' durations over-educated (i.e., the wage penalty is constant over time spent over-educated).
[60] Similar to the cross-sectional matching, these estimates are derived using quasi-exact matching on highest education level and gender, the logit models contain extensive controls for individual characteristics—measures for human capital, recent labour market experiences, demographic characteristics and family background (see Notes to Table 5.3 for exact list of variables contained in these categories)—and interaction terms aimed at satisfying the balancing tests, virtually all the over-educated are successfully matched (i.e., less than 1 per cent cannot be matched) and adequate numbers of well-matched individuals are used in the matching. Unlike the analyses for estimates in panel B, quasi-exact matching on highest education and gender is possible here because within the 20 groups—highest education (10) x gender (2)—there are sufficient numbers of well-matched individuals to match to the over-educated (i.e., because a small *treatment group* is being matched to a large *control group*). Logit model results with the chosen specifications, graphs establishing sufficient overlap exists in the propensity scores of the over-educated and well-matched and results indicating these estimates are not overly reliant on small numbers of well-matched individuals are not presented, but are available from the author on request.
[61] Results from these balancing tests are not presented, but are available from the author on request.

identify causal effects. That is, panel A contains the estimates reliant on the selection on observables assumption, while panel B contains those that allow for selection on unobservables.[62]

Table 5.8: Summary—Over-education wage penalty estimates (%)

Empirical estimators	Males		Females	
	Over-educ. effect	*N*	Over-educ. effect	*N*
Difference in (mean) real hourly wages	-18.3**	*21,587*	-15.9**	*22,014*
A. Selection on observables				
Regression estimators				
Pooled OLS	-15.1**	*20,288*	-16.3**	*20,806*
Cross-sectional matching estimators				
Nearest neighbour matching	-16.0**	*6,624*	-17.3**	*8,030*
Kernel matching	-14.6**	*17,052*	-14.8**	*16,915*
Combined matching and regression estimators				
Nearest neighbour and pooled OLS	-14.9**	*7,868*	-17.6**	*9,970*
Kernel matching and pooled OLS	-13.6**	*16,938*	-16.0**	*16,776*
B. Selection on unobservables				
Regression estimators				
Fixed effects	-3.9**	*20,664*	-6.0**	*21,069*
First-differences	-4.3*	*13,986*	-6.2**	*14,006*
Difference-in-differences matching estimators				
Restricted sample: Well-matched at *t′*	-9.3*	*516*	-7.1*	*530*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTE: ** and * indicate statistical significance at the 1% and 5% levels.

The differences in mean wages suggest wage differentials of around 18 per cent for males and 16 per cent for females. The pooled OLS, cross-sectional matching and combined matching and regression estimators all produce similar wage penalty estimates; they indicate wage penalties of roughly 14 to 16 per cent for males and 15 to 18 per cent for females. Given this similarity and the advantages of the latter estimators, the empirical evidence indicates that these estimates are not being distorted by the linear functional form assumption associated with regression analyses, extrapolation to outliers in the data or any discrepancies in the (observed) characteristics of the matched samples of over-educated and well-matched individuals. A priori, however, the pooled OLS estimates are expected to be upwardly biased—as the inability to account for individuals' innate abilities in the regression models means the conditional independence assumption is likely violated. The fixed effects and first-differences estimators produce much smaller wage penalty estimates of around 4 per cent for males and 6 per cent for females. Since these estimators control for the effects of unobserved individual heterogeneity, the estimates are considered empirical evidence that the pooled OLS, cross-sectional matching and combined matching and regression estimates are indeed upwardly

[62] More specifically, Table 5.8 contains: the differences in real hourly wages from Table 5.1; the pooled OLS estimates from specification (I) in Table 5.3; the cross-sectional matching estimates from Table 5.5; the combined matching and regression estimates from specification (I) in Table 5.6; the fixed effects and first-differences estimates from specification (II) in Table 5.4; and, the DID matching estimates (nearest neighbour) from panel C in Table 5.7.

biased. The fixed effects and first-differences estimates, however, are unlikely to represent causal effects because the individuals who identify them—those who move between the over-educated and well-matched states—are not representative of all over-educated individuals (i.e., due to selection effects in the variation in over-education status). Based on the differences in observed characteristics, as discussed in Sections 5.4.1 and 5.4.2, these estimates are expected to be downwardly biased, particularly for males.

Finally, there are the DID matching estimates. The DID matching estimator also controls for the effects of unobserved individual heterogeneity—and does so on the basis of considerably weaker assumptions—and the sample used to identify the estimates is approximately representative of all over-educated individuals (or, at least, more representative than the sample used to identify the fixed effects and first-differences estimates).[63] Moreover, these estimates lie between the other, biased estimates: they are significantly smaller than the (upwardly biased) pooled OLS, cross-sectional matching and combined matching and regression estimates, and larger than the (downwardly biased) fixed effects and first-differences estimates (with, as expected, this latter difference more pronounced for males). Ultimately, based on all of this empirical evidence, it is reasonable to conclude that the DID matching estimates can be interpreted as causal effects. Over-education, therefore, is found to have a negative, statistically significant effect on individuals' wages. And the size of the over-education wage penalty is, on average, approximately 9 per cent for males and 7 per cent for females. This means the pooled OLS estimates, which are similar to those derived in the over-education literature, contain considerable bias: controlling for the effects of unobserved individual heterogeneity reduces the size of the wage penalty by roughly 40 per cent for males and 55 per cent for females. Hence, the vast majority of previous studies have significantly over-estimated the size of the over-education wage penalty.

5.6 Discussion and conclusion

The aim of this chapter was to empirically test the validity of the method used to identify over-education. It was assumed wages reflect individuals' labour productivity achieved in the workplace and so the relationship between over-education and wages was examined, whereby evidence of an over-education wage penalty would validate the identification method. A series of empirical estimators was used to estimate the effect of over-education, with the aim being to establish estimates that can be interpreted as causal effects. Ultimately, over-education was found to have a

[63] Given its use of matching, the DID matching estimator also has the advantage of not requiring the linear functional form assumption (i.e., flexible parametric specifications are used to model non-random selection into over-education, weights are defined to ensure estimates represent the mean effect among the over-educated and sample restrictions are imposed to ensure the CSA is satisfied). Although, since a restricted sample is examined—recent entrants to over-education—it must be assumed that the over-education wage penalty does not vary by individuals' durations over-educated.

negative, statistically significant effect on individuals' wages: being over-educated *causes* an individual's wage to be lower than if they were instead well-matched. And the size of this over-education wage penalty is, on average, approximately 9 per cent for males and 7 per cent for females. This chapter, therefore, has found that the individuals identified as over-educated do indeed have human capital from education that is being under-utilised in their current job. Thus, the empirical evidence validates over-education as a labour market phenomenon—it is not merely a statistical artefact—and validates the method used to identify over-education. This means instances of over-education, as presented in Chapter 4, represent a source of potential gains for the economy: they signify an under-utilisation of the human capital available in the workforce, whereby having these over-educated individuals instead well-matched would result in higher wages for the individuals and then higher productivity levels, economic growth rates and living standards.

In the over-education literature, few studies have empirically tested the validity of over-education and its identification, and, due to endogeneity bias and data limitations associated with previous estimates, there is no clear evidence on whether over-education has a causal effect on individuals' wages. This chapter, therefore, has made several contributions to the literature. Specifically, it has provided empirical evidence that validates the existence of over-education in labour markets. It has provided over-education wage penalty estimates that can be interpreted as causal effects (i.e., found over-education has a negative causal effect on wages). And, moreover, it has found that pooled OLS estimates, as reported in most previous studies, significantly over-estimate the size of the over-education wage penalty—this (upward) bias is estimated at around 40 per cent for males and 55 per cent for females (and these may actually be conservative estimates of the bias in previous studies because this chapter has also improved on the previous pooled OLS estimates by controlling for a more extensive array of individual characteristics, which thereby reduces the potential for endogeneity bias in the pooled OLS estimates derived here).

The empirical test performed in this chapter has one key limitation: the estimates of the effect of over-education that have been derived represent merely the *average* effect among over-educated individuals. In reality, the effect is likely to vary across individuals. For instance, it may vary by individuals' highest education level, duration over-educated or severity of over-education (i.e., the difference between completed and required education levels). Hence, as discussed in Section 3.3.1, the evidence that there is, on average, a statistically significant over-education wage penalty does not necessarily guarantee that all the individuals identified as over-educated have under-utilised human capital. Examining the heterogeneity in the over-education wage penalty is a suitable avenue for future research; it is considered in Chapter 6, where the wage penalty is estimated by individuals' duration over-educated.

Chapter 6

Dynamics of over-education

6.1 Introduction

Chapters 4 and 5 have established there is evidence of over-education in Australian labour markets, with estimates indicating that its incidence has been relatively constant over the 2001–2008 period. What remains unclear is whether it is the same individuals over-educated in each year or whether over-education is actually a result of the inherently dynamic nature of modern labour markets. Modern labour markets are complex and continuously evolving: there are always individuals investing in human capital, moving in and out of the labour force and moving between jobs, and firms are always creating, adjusting and discontinuing jobs. And this may result in the individual-job mismatches that lead to over-education. But, with time, it may also lead to their resolution as either individuals, in seeking to maximise their utility (or wages), move to jobs that fully utilise their human capital, or firms, in seeking to maximise their profits, adjust jobs to fully utilise individuals' human capital. Hence, instances of over-education may be short-term disequilibria. If the time spent over-educated has no enduring effects, over-education could then be regarded as merely a by-product of adjustment processes in dynamic and well-functioning labour markets. Alternatively, if over-education is persistent for individuals or has enduring effects, then it represents more serious labour market failures. This chapter, therefore, considers the research question: *are instances of over-education short-term labour market disequilibria that have no enduring effects?*

Empirical evidence is derived from two sets of analyses. The first considers whether over-education is merely short-term disequilibria. This requires evidence on the two ways in which instances of over-education can be resolved: individuals changing jobs and firms adjusting jobs. Since HILDA Survey data capture individuals' movements between jobs, evidence of resolutions from the first source is derived via examination of the changes in individuals' over-education statuses over time.[1] Hence, descriptive statistics on individuals' transitions to and from over-education and their durations over-educated are calculated. Such analyses, however, may not capture evidence of resolutions from the second source because individuals may not report adjustments to their jobs

[1] Recall from Chapter 3, HILDA Survey data actually capture changes in occupations, not jobs. But such changes are sufficient for examining resolutions of over-education because occupations are examined using ANZSCO at the 4-digit level and any job changes occurring within these occupation categories (which cannot be identified in the data) are unlikely to result in a change in over-education status as virtually all jobs in the same category have the same required skill level.

made by their employer as a change of occupation. That is, if the adjustments to individuals' jobs that lead to the full utilisation of their human capital cannot be identified in the data, then affected individual-job matches would appear unchanged and the individuals, erroneously, remain identified as over-educated. In order to ascertain whether such resolutions occur, the relationship between duration over-educated and wages is examined. Specifically, the over-education wage penalty is estimated by individuals' duration over-educated, and evidence the wage penalty disappears with time spent over-educated is considered indicative of over-education being resolved by firms adjusting jobs. That is, since it is again assumed wages reflect the labour productivity achieved in the workplace, evidence the wage penalty disappears is considered indicative of individuals who appear to have been over-educated for extended periods no longer actually having under-utilised human capital—instead, their human capital is utilised to the same extent as (otherwise identical) well-matched individuals—and, presumably, these instances of over-education were resolved because firms adjusted the jobs to fully utilise their human capital. As in Chapter 5, the fixed effects estimator is used in an attempt to derive wage penalty estimates that can be interpreted as causal effects. Finding individuals are only temporarily identified as over-educated or the over-education wage penalty disappears over time is, therefore, assumed to be evidence that instances of over-education are short-term labour market disequilibria.

The second set of analyses considers whether over-education has enduring effects. Since over-education represents individuals with under-utilised human capital, and skill atrophy can result from periods in which skills are insufficiently used, the most likely enduring effect is human capital depreciation. Hence, this chapter focuses on the possibility that over-education leads to human capital depreciation. Two empirical tests are performed. One examines whether being over-educated in the past increases the likelihood of being over-educated again in the future—it tests for state dependence in over-education. Specifically, the effect that being over-educated in the previous year has on an individual's current likelihood of being over-educated is estimated, with evidence that it increases the likelihood considered indicative of state dependence in over-education. Evidence of state dependence is sought as it is commonly argued to be evidence of human capital depreciation (see, for example, Pissarides (1992), Arulampalam, Booth and Taylor (2000) and Stewart (2007)). Since state dependence is defined in terms of a causal relationship, it is important the estimates represent causal effects. This requires an econometric model (and empirical estimators) capable of separately identifying the causal effects of prior over-education and (observable and unobservable) individual characteristics. Dynamic panel probit models (and the Heckman (1981c), Orme (1997) and Wooldridge (2005) estimators) are therefore used to derive the estimates.

The other empirical test for human capital depreciation examines the relationship between prior over-education and the wages of currently well-matched individuals. Specifically, the effect that

being over-educated in the previous year has on their current wages is estimated, with evidence that it reduces wages considered indicative of individuals having reduced labour productivity due to human capital depreciation. That is, since it is assumed wages reflect labour productivity and the sample is restricted to currently well-matched individuals, evidence the previously over-educated have reduced wages, compared to (otherwise identical) individuals who were well-matched, is considered indicative of these individuals having reduced labour productivity due to human capital depreciation (which presumably occurred while they were over-educated).[2] For this analysis, the difference-in-differences matching estimator is used in an attempt to derive estimates that can be interpreted as causal effects. Finding state dependence in over-education or prior over-education reduces the wages of well-matched individuals is assumed to be evidence that over-education leads to human capital depreciation, and, therefore, evidence that instances of over-education have enduring effects.[3]

This chapter proceeds as follows. Section 6.2 briefly recaps existing evidence on the persistence of over-education and the possibility over-education leads to human capital depreciation. Section 6.3 presents evidence on persistence, with descriptive statistics on transition rates and durations over-educated and then wage penalty estimates by duration over-educated. Section 6.4 discusses state dependence and outlines dynamic panel probit models and the empirical estimators used to derive estimates, and then presents evidence regarding state dependence in over-education. Section 6.5 examines the relationship between prior over-education and the wages of well-matched individuals. Section 6.6 concludes the chapter.

6.2 Existing evidence

As discussed in Chapter 2, there is little empirical evidence regarding the dynamics of over-education. The evidence on persistence is predominantly based on simple cross-tabulations of individuals' over-education status between two points in time. Such evidence is mixed: findings from some studies suggest over-education is highly persistent, while the results of others suggest much less persistence. And, more importantly, it does not accurately measure the factor most relevant to persistence: the

[2] The examination of individuals' wages is the most frequently used approach for establishing empirical evidence of human capital depreciation (De Grip and Van Loo, 2002; De Grip, 2006) (see, for example, Mincer and Polachek (1974), Mincer and Ofek (1982), Albrecht et al. (1999), Kunze (2002), Arrazola and De Hevia (2004) and Görlich and De Grip (2009)).

[3] Evidence of human capital depreciation and a disappearing wage penalty would also have implications for the identification and interpretation of over-education. If individuals' human capital depreciates while over-educated, then it is possible some individuals no longer possess the human capital from their education that is, or was previously, under-utilised in their current job; thus, they are incorrectly identified as over-educated and would, at least notionally, be better characterised as well-matched. Alternatively, if no human capital depreciation occurs and the wage penalty disappears with time spent over-educated, then it is possible some individuals are no longer under-utilising their human capital from education and would, as a result, also be better characterised as well-matched. Such evidence, therefore, would mean not all instances of over-education identified in data necessarily represent individuals with under-utilised human capital. But, to be clear, in the first instance previously over-educated individuals become 'well-matched' due to a loss of human capital, while in the second instance it is a greater utilisation of their human capital that leads to the change. Both cases would also then affect the costs of over-education.

durations of individuals' spells over-educated. Nevertheless, the existing evidence appears to indicate that over-education is not an entirely temporary phenomenon, particularly among recent labour market entrants. Meanwhile, only two studies—Rubb (2006) and De Grip et al. (2008)—have examined the possibility that over-education leads to human capital depreciation. Both found evidence to support such a link, but the relevant estimates in Rubb (2006) are likely biased and De Grip et al. (2008) found only weak evidence of human capital depreciation, though this is most likely because they examined individuals' general aptitudes (or cognitive abilities) rather than job-specific human capital (where the former may be less likely to atrophy while individuals are over-educated).

This chapter contributes in several ways to the over-education literature. First, it adds to the limited empirical evidence on the persistence of over-education; most notably, it provides some of the first evidence on individuals' durations over-educated. It is the first study to examine dynamics of the over-education wage penalty. Specifically, it is the first to consider whether the wage penalty varies by individuals' durations over-educated and whether prior over-education affects the wages of currently well-matched individuals. It is also the first study to test for state dependence in over-education, with the resultant evidence (along with the evidence regarding the wage effects of prior over-education) contributing to the limited empirical evidence on the potential link between over-education and human capital depreciation.

6.3 Persistence of over-education

In deriving empirical evidence on the persistence of over-education, the two ways in which instances of over-education can be resolved—individuals changing jobs and firms adjusting jobs—must both be considered. This is done in the two sub-sections below. Specifically, evidence on individuals' transition rates and durations over-educated capture resolutions from the first source, while estimates of the over-education wage penalty by individuals' durations over-educated aim to provide evidence of resolutions from the second source.

6.3.1 Transition rates and durations over-educated

Based on analysis of adjacent waves of the data, Table 6.1 presents rates of continuation (column [2]), inflow ([3]) and outflow ([4]) for over-education over one-year intervals; meanwhile, column [1] contains the average incidence of over-education over the 2001–2008 period. The figures indicate roughly 85 per cent of males over-educated in a given year are also over-educated the following year, and similarly for 87 per cent of over-educated females. Thus, it appears only 15 per cent of over-educated males and 13 per cent of over-educated females exit the state each year. There is also a

small inflow to over-education, with around 3 per cent of previously well-matched (or under-educated) males and females entering the state each year.[4] The vast majority of these transitions occur because individuals change occupation (see Table A6.1.2 in Appendix 6.1 for the sources of over-education entries and exits over the one-year intervals). Table 6.1 also presents elementary estimates of the state dependence in over-education: the differences between the rates of continuation and entry (in column [5]). If there is no state dependence (i.e., the likelihood of being over-educated is independent of previous over-education status) and all individuals are equally likely to be over-educated, then the continuation and entry rates would be identical and these estimates would equal zero. This is not the case, and instead the differences are in excess of 80 percentage points.[5] These estimates and the state dependence in over-education are further discussed in Section 6.4.

Table 6.1: Over-education transition rates over one-year intervals by gender

	[1] *Over-educated*	[2] *Remain over-educated*	[3] *Inflow*	[4] *Outflow*	[5] *State dependence*	[6] *Ratio*
	$\Pr(O_t=1)$	$\Pr(O_t=1 \mid O_{t-1}=1)$	$\Pr(O_t=1 \mid O_{t-1}=0)$	$\Pr(O_t=0 \mid O_{t-1}=1)$	([2]–[3])	([2]/[3])
Males	0.161	0.845	0.027	0.155	0.818	31.3
Females	0.200	0.865	0.028	0.135	0.837	30.9
All	0.180	0.856	0.028	0.144	0.828	30.6

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: Figures are based on pooled two-year panels of adjacent waves of the data.
Sample examined contains 33,996 individuals (17,359 males and 16,637 females), with 6,109 over-educated at *t*, 6,242 over-educated at *t-1* and 27,754 not over-educated (i.e., well-matched or under-educated) at *t-1*.

The persistence of over-education can also be considered by examining the total number of periods (or waves) in which individuals are observed over-educated during the eight-year period. Table A6.1.3 in Appendix 6.1 presents such analyses. While these results are informative—for instance, they indicate roughly 9 per cent of the males and 10 per cent of the females employed (and in-sample) in each year are over-educated in all eight years—it is arguably more insightful to consider them in conjunction with the cross-sectional incidences of over-education. Such analyses provide a clearer indication of the proportion of over-education in each year that represents persistent over-education. Table 6.2 presents these results. Based on the assumption that being over-educated in

[4] Table A6.1.1 in Appendix 6.1 presents a more detailed analysis of such transitions: for each two-year panel, it tracks individuals' transitions between the over-educated, under-educated, well-matched, not employed and not in-sample states, where 'not in-sample' means individuals are either 65 years of age or older, a full-time student, self-employed or a survey non-respondent in a particular year. Rates of continuation (or persistence) in over-education are significantly lower in this supplementary table—they are between 61 and 66 per cent, compared to the 85 and 87 per cent in Table 6.1—and this is because the not employed and not in-sample individuals are included in the analysis. Table A6.1.1 also indicates that roughly 3 per cent of previously not employed individuals enter the over-educated state each year.

[5] That is, being over-educated in the previous year increases the likelihood an individual is currently over-educated by roughly 80 percentage points. Alternatively, column [6] of Table 6.1 presents the ratios of the probabilities: these indicate that individuals over-educated in the previous year (compared to those who were not) are roughly 31 times more likely to be currently over-educated.

seven or eight of the years corresponds to persistent over-education, the figures indicate a sizeable proportion of over-education in each year is accounted for by persistent over-education: it accounts for approximately 40 per cent of the over-education among males and 35 per cent of over-education among females in any year. There are also large proportions over-educated in five or six of the years; arguably these are also representative of persistent over-education (and, if considered as such, would mean persistent over-education accounts for roughly 65 per cent of the over-education among males and 60 per cent of over-education among females). At the other extreme, between 10 and 25 per cent of over-education is temporary (i.e., accounted for by individuals over-educated in only one or two years).[6] With such distinctions between *persistent* and *temporary* over-education, there is also the in-between group of individuals who are considered neither persistently nor temporarily over-educated; in this case, those over-educated in three or four years (or three to six years depending on the assumption as to what represents persistent over-education).

As previously discussed, the factor most relevant to the persistence of over-education is the durations of individuals' spells over-educated. While the HILDA Survey data do not contain a measure that directly captures this factor, it can be approximated using the measure for individuals' tenure in their current occupation. Such an approximation, however, has two limitations. First, it must be assumed that individuals have been over-educated throughout their time in the occupation (i.e., they did not become over-educated during the period due to the completion of a higher level qualification). If not the case, these occupation tenures would over-estimate individuals' durations over-educated. Since Table A6.1.2 (in Appendix 6.1) indicates the vast majority of entries to over-education occur because individuals change occupation (rather than increase education), this assumption appears reasonable and unlikely to impart significant bias to the resultant evidence. Second, the durations represent individuals' time spent over-educated in their current occupation (at the time of the interview), but not necessarily the *total* time spent over-educated because they refer to ongoing spells of over-education and do not capture any time spent over-educated in other occupations. This means the occupation tenures are likely to under-estimate the durations of individuals' spells over-educated.[7]

---

[6] Survival functions for recently over-educated individuals are another means by which persistence can be considered; these report the proportions who remain over-educated by the number of years since they were first observed over-educated. Such figures are presented in Figure A6.1.1 in Appendix 6.1. Results indicate between 30 and 50 per cent of individuals exit over-education within approximately one year (though this includes individuals who are no longer in-sample), while roughly 20 per cent of males and 15 per cent of females remain over-educated seven years later. The limitation of this analysis, however, is that it only considers recent entrants (and not all the over-educated individuals in the data).

[7] Also, given the wording of the HILDA Survey questions, these occupation tenures do not necessarily represent continuous employment in the occupation and, as a result, a further limitation of this approximation is that the durations may include previous spells of over-education.

Table 6.2: Cross-sectional incidences of over-education by total number of periods observed over-educated, gender and year—Individuals in balanced panel sample (%)

Over-educated at t Total number of periods (or waves) observed over-educated	Year (wave) which corresponds to *t*							
	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Males								
Over-educated at t	*16.2*	*18.3*	*19.3*	*18.5*	*18.2*	*19.4*	*18.7*	*19.6*
8	32.2	31.0	28.1	29.9	29.4	29.0	28.9	27.0
7	8.2	10.4	11.6	11.9	11.9	11.8	11.7	9.1
6	9.5	12.7	15.6	16.1	13.2	15.7	13.1	8.9
5	8.9	9.0	9.9	13.8	14.7	10.1	7.7	7.4
4	5.9	6.3	6.1	8.2	9.4	8.9	9.7	7.0
3	10.5	11.0	14.7	10.6	10.9	13.8	12.7	13.2
2	11.9	12.0	7.4	5.6	4.3	5.6	9.7	10.8
1	13.0	7.7	6.6	3.9	6.1	5.1	6.5	16.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	298	302	322	308	305	320	317	335
Females								
Over-educated at t	*21.5*	*21.3*	*22.5*	*22.4*	*23.4*	*24.1*	*24.1*	*23.8*
8	26.2	26.6	23.5	22.1	22.0	20.6	21.2	21.0
7	10.6	12.9	11.9	10.1	10.7	9.5	9.8	8.0
6	10.3	13.9	16.2	15.7	15.7	15.9	13.2	13.1
5	10.5	12.0	12.8	18.4	19.3	14.5	11.8	11.3
4	9.6	7.8	9.2	13.6	12.5	12.5	12.6	9.7
3	11.5	11.3	12.4	9.0	7.8	13.7	12.3	13.6
2	9.2	9.9	6.0	5.4	5.9	7.3	13.6	12.1
1	12.3	5.6	8.1	5.7	6.2	6.1	5.4	11.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	390	373	417	414	430	458	449	459

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on the sample of individuals in the balanced panel; N refers to the number over-educated at *t*.
Italicised figures are the estimated incidences of over-education reported in Chapter 4.
The remaining figures are proportions that sum to 100.0 for (sub-sections in) each column and are weighted using longitudinal population weights (for the W1–W8 balanced panel) to make them representative of the Australian population of individuals over-educated in each year.

Table 6.3 presents the evidence on duration over-educated for the individuals over-educated in each year. Similar to the results in Table 6.2, these figures indicate a sizeable amount of persistence in over-education: approximately 20 per cent of over-educated males and females have been over-educated for ten years or more, while a further 10 per cent have been over-educated for six to nine years. At the other extreme, roughly 25 per cent have been over-educated for less than a year, though, as discussed above, this likely overstates the extent to which over-education is temporary as the durations refer to ongoing spells (and do not capture any time spent over-educated in other occupations). The results in Table 6.3 also indicate that the currently over-educated individuals have, on average, been over-educated for between five and six years.

Table 6.3: Duration over-educated (at time *t*, not total) by gender and year—Individuals over-educated at *t* (%)

	Year (wave) which corresponds to *t*							
	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Males								
Less than 1 year	24.2	27.0	24.6	22.8	24.8	23.9	23.6	22.4
1–2 years	25.8	22.6	21.3	21.1	19.0	19.3	20.3	25.6
3–5 years	20.4	20.0	25.4	24.8	25.0	21.8	20.6	19.0
6–9 years	11.8	11.8	10.0	8.6	10.1	14.9	11.8	13.2
10 or more years	17.8	18.7	18.7	22.8	21.0	20.2	23.8	19.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	490	519	540	510	517	567	563	596
Average number of years	5.32	5.55	5.38	5.93	5.88	5.87	6.08	5.91
Females								
Less than 1 year	25.1	24.5	26.4	24.1	26.2	26.6	25.0	26.6
1–2 years	21.8	21.3	22.3	23.6	22.2	20.5	18.5	19.3
3–5 years	21.6	22.3	19.9	22.2	20.6	21.2	22.3	22.3
6–9 years	14.0	14.6	14.8	13.9	12.8	10.0	10.5	12.1
10 or more years	17.4	17.3	16.6	16.2	18.3	21.7	23.7	19.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	640	581	616	619	704	742	728	732
Average number of years	4.91	5.05	4.94	5.06	5.06	5.45	5.72	5.44

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for (sub-sections in) each column (except the averages) and are weighted using cross-sectional populations weights to make them representative of the Australian population of individuals over-educated in each year.

Duration over-educated is approximated using the measure for individuals' tenure in their current occupation. This approximation has two limitations: it must be assumed individuals have been over-educated throughout their time in the occupation (i.e., did not become over-educated during the period due to the completion of a higher level qualification); and, the durations represent individuals' time spent over-educated in their current occupation (at time of interview at *t*), but not necessarily the *total* time over-educated as they refer to ongoing spells of over-education and do not capture any time spent over-educated in other occupations.

All of the descriptive statistics discussed above have clearly indicated that it is not the case that individuals are only temporarily identified as over-educated. Instead, it appears over-education is persistent for a sizeable group of individuals: generally, 30 to 40 per cent of the over-education in each year corresponds to individuals who are persistently over-educated, with the majority of these individuals seemingly over-educated for in excess of ten years. The issue of whether these persistently over-educated individuals have under-utilised human capital—that is, whether or not they actually remain over-educated after such extended periods—is considered in the following sub-section.

6.3.2 Wage penalty by duration over-educated

Individuals identified as persistently over-educated may no longer have under-utilised human capital because firms can adjust jobs over time to enable a greater utilisation of individuals' human capital and because individuals may not report such adjustments as a change in occupation. Hence, some individuals may erroneously remain identified as over-educated. If indeed the case, these individuals would no longer incur wage penalties. The over-education wage penalty is, therefore, estimated by

individuals' duration over-educated in order to derive evidence of such resolutions. Specifically, evidence the wage penalty disappears with time spent over-educated is considered indicative of instances of over-education being resolved by firms adjusting jobs.

Econometric model and estimators

As in Chapter 5, an important requirement for the above conclusion is that the wage penalty estimates represent causal effects, not just conditional associations. A model similar to [5.3]—an ORU earnings function—is therefore estimated using the fixed effects estimator (as defined in Section 5.3.2). In particular, the measure for duration over-educated (discussed above) is used to replace the O_{it} indicator in [5.3] with either duration over-educated (in years) (along with its squared-term to capture any non-linear effects of duration over-educated) or a set of indicators for duration over-educated (where the categories considered are: *less than 1 year*; *1–2 years*; *3–5 years*; *6–9 years*; and, *10 or more years*).[8] Apart from these differences in the model specification, the empirical analyses performed here essentially mirror those conducted in Chapter 5. The same sample is examined; namely, unbalanced panels of over-educated and well-matched individuals. Analyses again examine real hourly wages (expressed in 2001-dollars) and estimate effects separately for males and females. Robust (panel-corrected) standard errors are calculated to ensure valid statistical inference. The robustness of the results is considered by replicating the analyses using balanced panels. And the models contain the same extensive set of controls for individual characteristics; that is, human capital levels, recent labour market experiences, demographic characteristics (excluding ethnicity) and year dummies.[9, 10]

For the specification with indicators for duration over-educated, several Wald tests on the statistical significance of the estimated coefficients are performed. The first tests whether the coefficients for the duration over-educated indicators are jointly statistically significant, while the second tests whether these coefficients are equal to one another. Additional Wald tests then consider equality for all possible pairs of these coefficients—that is, whether the coefficient for *Over-educated: Less than 1 year* is equal to the coefficient for *Over-educated: 1–2 years*, and so on for all remaining pairs (*Over-educated: Less than 1 year* vs. *Over-educated: 3–5 years*, *Over-educated: Less than 1 year* vs. *Over-educated: 6–9 years*, *Over-educated: Less than 1 year* vs. *Over-educated: 10+ years*, *Over-educated: 1–2 years* vs. *Over-educated: 3–5 years*, etc).

[8] These indicators are effectively interactions between the indicator for being over-educated at *t* (i.e., O_{it}) and the measure for duration over-educated at *t* (i.e., duration of current spell of over-education).
[9] The exact list of variables contained in these categories is reported in the Notes to Table 5.3. For further details on the explanatory variables (and the dependent variable, real hourly wage) see Appendix 5.1.
[10] Pooled OLS estimates are also derived. But since, as established in Chapter 5, such estimates are likely upwardly biased (i.e., not representative of causal effects), they are presented in an appendix and only briefly considered.

Using the fixed effects estimator to derive wage penalty estimates, however, is seemingly at odds with previous findings in this study. Recall, Chapter 5 concluded that the difference-in-differences (DID) matching estimator produces the preferred wage penalty estimates (i.e., those most likely to represent causal effects), whereas the fixed effects estimates are likely downwardly biased. Thus, the DID matching estimator might also be preferred when seeking wage penalty estimates by individuals' durations over-educated. Such estimates, however, cannot be derived. This is because Chapter 5 also concluded that quasi-exact matching on individuals' over-education status at *t–1* is necessary for the DID matching estimator to produce valid causal effects estimates, and that, in fact, valid estimates can only be produced for individuals well-matched at *t–1*. Given such a sample restriction, the DID matching estimator cannot identify wage penalty estimates by duration over-educated because all of the individuals examined—those well-matched at *t–1*—would, at least notionally, have been over-educated for less than one year. As a result, the next-best estimator, preferably one designed to control for unobserved individual heterogeneity, must be used. It is for this reason the fixed effects estimator is used here.

Empirical results

Prior to examining the fixed effects estimates, descriptive statistics on individuals' wages and duration over-educated are considered. Table 6.4 presents means and standard deviations for the hourly wages of the over-educated—by duration over-educated—and well-matched individuals, along with the differences in their mean wages. These differences are elementary estimates of the over-education wage penalty by duration over-educated, and they are all statistically significant at the 1 per cent level, except for males over-educated ten or more years and females over-educated between six and nine years. The figures indicate the wage penalty declines with time spent over-educated. For males, it declines from roughly 29 per cent in the initial year over-educated to essentially zero after ten years over-educated. Meanwhile, for females, the wage penalty is initially around 23 per cent and then remains significant at 9 per cent after ten years over-educated.[11] Such evidence, however, confounds the effects of duration over-educated with the effects that age and labour market experience (and other human capital factors) can have on individuals' wages (i.e., these wage penalty estimates are not derived from a comparison with otherwise identical individuals).

[11] Recall from Table 5.1 in Chapter 5, the overall differences are roughly 18 per cent for males and 16 per cent for females.

Table 6.4: Average wages by duration over-educated and gender

	Males				Females			
	Mean	*Difference*		*N*	Mean	*Difference*		*N*
		($)	(%)			($)	(%)	
Well-matched	22.52			*17,294*	19.60			*16,659*
	(15.57)				(13.28)			
Over-educated: Less than 1 year	16.02	-6.49	-28.8	*1,088*	15.08	-4.52	-23.1	*1,392*
	(9.31)				(12.32)			
Over-educated: 1–2 years	17.11	-5.41	-24.0	*927*	15.49	-4.10	-20.9	*1,141*
	(11.11)				(7.14)			
Over-educated: 3–5 years	17.31	-5.21	-23.1	*925*	16.65	-2.94	-15.0	*1,150*
	(8.52)				(12.21)			
Over-educated: 6–9 years	18.91	-3.60	-16.0	*484*	18.84	-0.76	-3.9	*655*
	(8.03)				(18.43)			
Over-educated: 10+ years	22.76	0.25	1.1	*867*	17.78	-1.81	-9.3	*1,011*
	(11.79)				(10.87)			
N				*21,585*				*22,008*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Real hourly wages examined; see Table A5.1.1 in Appendix 5.1 for details on the derivation of real hourly wages. Standard deviations reported in parentheses.
All differences are statistically significant at the 1% level, except that for males over-educated 10+ years and females over-educated 6–9 years.

Since the fixed effects estimates for the duration over-educated indicators are identified by the individuals whose values for the indicators change over time, Table A6.2.1 in Appendix 6.2 presents information on the amount of variation in these indicators. Similar to the analyses in Chapter 5, the results indicate that each of these key indicators contains limited *within* variation. A relevant concern then, for each indicator, is whether the individuals whose values change over time (i.e., those who move between the various durations over-educated or who move between the particular duration over-educated and the well-matched state) are representative of all individuals over-educated for that particular duration. Recall from Section 5.3, this representativeness affects whether the fixed effects estimates can be interpreted as causal effects. Hence, the representativeness is empirically tested using the (mean) observable characteristics of the groups (i.e., t-tests similar to those discussed in Section 5.4.1 are performed).[12] For each duration over-educated indicator, the results indicate the sub-samples whose values change are not entirely representative. In particular, these sub-samples typically contain higher proportions of younger individuals (aged 15 to 34 years) and more highly educated individuals (i.e., those with a Bachelor Degree or higher), with the differences more pronounced among males. As argued in Section 5.4.2, such differences mean the fixed effects estimates are likely to be downwardly biased.

[12] These t-tests are conducted using an extensive set of individual characteristics and they determine, for each characteristic, whether there are statistically significant differences in the means of the two groups being considered (e.g., individuals over-educated for less than 1 year who are later (or previously) observed well-matched or over-educated for more than 1 year vs. all individuals observed over-educated for less than 1 year). Tests are performed separately for each duration over-educated indicator—*Over-educated: Less than 1 year*, *Over-educated: 1–2 years*, *Over-educated: 3–5 years*, *Over-educated: 6–9 years* and *Over-educated: 10+ years*—and conducted by gender. The results are not presented, but are available from the author on request.

Table 6.5 presents the fixed effects estimates of the over-education wage penalty by duration over-educated. Specifically, estimates from the specification using duration over-educated (in years) are presented in column (I) and those using the duration over-educated indicators are in column (II), and then the estimates from these specifications using balanced panels are in columns (III) and (IV). Table A6.2.2 in Appendix 6.2 presents the complete set of results for the models in (I) and (II).[13] Results in (I) and (III) indicate the wage penalty increases with time spent over-educated, though at a decreasing rate—that is, a non-linear relationship appears to exist between the wage penalty and duration over-educated. These estimates, however, are not statistically significant for females. Contrastingly, the results from the more flexible specification in (II) indicate the wage penalty is relatively constant: it is approximately 3 to 5 per cent for males and 4 to 6 per cent for females. And the Wald tests for joint significance indicate the coefficients for the duration over-educated indicators are jointly statistically significant, while the tests for equality cannot reject the possibility that these coefficients are equal to one another.[14]

Results from the sensitivity analyses in (IV), however, are not entirely consistent with the evidence of a relatively constant wage penalty. For males, the wage penalty estimates are larger and exhibit greater variation across duration over-educated—ranging from 3 to 11 per cent—and the Wald tests for equality indicate statistically significant differences exist between the coefficients. These estimates also suggest the wage penalty may increase with time spent over-educated. This is contrary to the evidence in (II), but, importantly for the interpretation of the results derived in this sub-section, it is not evidence that the wage penalty disappears with time over-educated. For females, the wage penalty estimates are smaller—ranging from 3 to 5 per cent—and they remain consistent with the conclusion that the wage penalty is relatively constant. These estimates, however, are not statistically significant. Overall, the differences between the results in (II) and (IV) arise because the balanced panels contain far fewer individuals who experience variation in the duration over-educated indicators. For this reason, results in (II) are considered the preferred estimates of the over-education wage penalty by duration over-educated.

[13] Complete results for models (III) and (IV) are not presented, but are available from the author on request.
[14] The additional Wald tests of equality for all possible pairs of coefficients also suggest a relatively constant wage penalty, with a statistically significant difference only found to exist between the coefficients for *Over-educated: Less than 1 year* and *Over-educated: 3–5 years* for females.

Table 6.5: Over-education wage penalty by duration over-educated—ORU earnings functions, fixed effects estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
Duration over-educated	-0.008**		-0.012**		-0.004		-0.005	
	(0.002)		(0.003)		(0.003)		(0.003)	
Duration over-educ.-squared	0.001**		0.001**		0.000		0.000	
	(0.000)		(0.000)		(0.000)		(0.000)	
Over-educated: Less than 1 year		-0.033		-0.071*		-0.078**		-0.029
		(0.018)		(0.036)		(0.018)		(0.029)
Over-educated: 1–2 years		-0.019		-0.035		-0.059**		-0.040
		(0.019)		(0.035)		(0.017)		(0.033)
Over-educated: 3–5 years		-0.043*		-0.034		-0.042*		-0.036
		(0.018)		(0.027)		(0.018)		(0.025)
Over-educated: 6–9 years		-0.051*		-0.103**		-0.031		-0.054
		(0.021)		(0.029)		(0.020)		(0.029)
Over-educated: 10+ years		-0.051*		-0.110**		-0.061**		-0.050
		(0.023)		(0.028)		(0.021)		(0.026)
Balanced panel sample	No	No	Yes	Yes	No	No	Yes	Yes
N	20,636	20,636	6,768	6,768	21,035	21,035	6,240	6,240
(No. individuals)	(5,323)	(5,323)	(846)	(846)	(5,449)	(5,449)	(780)	(780)
Within R-squared	0.0675	0.0676	0.0936	0.0951	0.0407	0.0426	0.0550	0.0554
Between R-squared	0.2036	0.2075	0.0749	0.0787	0.1512	0.1588	0.0003	0.0005
Overall R-squared	0.1598	0.1636	0.0675	0.0706	0.1059	0.1132	0.0006	0.0007
Wald test for joint significance of duration coefficients (H_0: All = 0) (p-value)		Reject H_0 (0.046)		Reject H_0 (0.000)		Reject H_0 (0.000)		Do not reject H_0 (0.377)
Wald test for equality of duration coefficients (H_0: All equal) (p-value)		Do not reject H_0 (0.534)		Reject H_0 (0.009)		Do not reject H_0 (0.197)		Do not reject H_0 (0.921)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *human capital measures*; *recent labour market experiences*; *demographic characteristics* (excluding ethnicity); and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories.

See Table A6.2.2 in Appendix 6.2 for the complete set of results for models (I) and (II).

Additional Wald tests for equality of duration coefficients indicate the following statistically significant differences between coefficients: *Less than 1 year* and *3–5 years* for females in (II); *3–5 years* and *6–9 years* for males in (IV); and, *3–5 years* and *10+ years* for males in (IV). Results from these Wald tests are not presented, but are available from the author on request.

These relatively constant wage penalty estimates—approximately 3 to 5 per cent for males and 4 to 6 per cent for females—are smaller than the preferred estimates in Chapter 5 (which are interpreted as representing causal effects), but similar to the fixed effects estimates that were derived. Recall from Section 5.5, the preferred wage penalty estimates in Chapter 5 are roughly 9 per cent for males and 7 per cent for females, while the fixed effects estimates are around 4 per cent for males and 6 per cent for females. This indicates therefore, as argued above, that the fixed effects estimates

of the wage penalty by duration over-educated are likely downwardly biased. But if it is assumed this bias is constant across duration over-educated (i.e., the size of the downward bias is approximately equal across the coefficients for the duration over-educated indicators), then it remains the case that the over-education wage penalty is relatively constant over time spent over-educated.[15] Regardless, the key result from these analyses is that there is no evidence the wage penalty disappears with time spent over-educated. This means there is no evidence of instances of over-education being resolved by firms' adjustments to jobs, and that individuals identified as persistently over-educated still have under-utilised human capital and, therefore, remain over-educated even after such extended periods.[16]

Based on the empirical evidence presented in this section, there are two clear findings: it is not the case that individuals are only temporarily identified as over-educated, and the over-education wage penalty does not disappear with time spent over-educated. Instead, it appears over-education is persistent for a sizeable group of individuals and the wage penalty is relatively constant over time spent over-educated. It has been found, therefore, that instances of over-education are not merely short-term labour market disequilibria.

6.4 State dependence in over-education

A state (or event) exhibits state dependence when experiencing the state in the past has a causal effect on the likelihood of experiencing it again in the future. Specifically, experiencing the state affects individuals' preferences, constraints and choices available during future decisions; past experience exerts a behavioural effect as an otherwise identical individual with no past experience would behave differently or face different opportunities (Heckman, 1981a). Such a causal relationship is typically referred to as *structural* or *true* state dependence (Heckman, 1981a; Heckman, 1981c; Stewart and Swaffield, 1999; Arulampalam et al., 2000; Stewart, 2007). In this study, finding such state dependence in over-education is assumed to be evidence that being over-educated leads to human capital depreciation. That is, since skill atrophy can result from periods in which skills are insufficiently used, it is assumed time spent over-educated leads to depreciation of individuals'

[15] The pooled OLS estimates, presented in Table A6.2.3 in Appendix 6.2, provide further evidence of a relatively constant wage penalty: the estimates (which are significantly larger than the fixed effects estimates as they are likely upwardly biased) are relatively constant at 15 to 20 per cent for males and 15 to 16 per cent for females, except among those over-educated for ten years or more as the wage penalty declines to roughly 11 per cent for males and increases to 21 per cent for females. This indicates the wage penalty may decline for males, but only marginally (and not to the point of insignificance).

[16] The results, of course, are not definitive evidence that no such resolutions occur. That is, since the derived estimates represent merely the *average* effects among the individuals in each duration over-educated category, the finding that for each category there is, on average, a statistically significant wage penalty does not necessarily guarantee that all the individuals in each category experience such a wage penalty. Hence, it is possible a small number of these individuals do not incur a wage penalty because their jobs were adjusted over time to ensure the full utilisation of their human capital (i.e., they are no longer actually over-educated).

(under-utilised) human capital and this then reduces their likelihood of obtaining a well-matched job. Being over-educated in the past, therefore, increases the likelihood of being over-educated again in the future.

Based on the elementary estimates presented in Section 6.3.1, there appears to be substantial state dependence in over-education: being over-educated in the previous year increases the likelihood an individual is currently over-educated by roughly 80 percentage points (or, compared to those not previously over-educated, it makes them roughly 31 times more likely to be currently over-educated). But, as highlighted by Heckman (1981a), such evidence—the observed persistence of a state in raw data—is not necessarily representative of the causal relationship that defines state dependence, and may instead be the result of individual heterogeneity. With individual heterogeneity, there are differences between individuals (either observable or unobservable) that affect the likelihood of experiencing the state and these differences (which are not affected by the experience of the state) persist over time. Hence, some individuals may be observed persistently over-educated because certain characteristics increase their likelihood of being over-educated and these characteristics persist over time. The failure to control for such characteristics in multivariate analyses would result in further (misleading) evidence that prior over-education significantly affects the likelihood of being over-educated again in the future; such an observed conditional relationship is typically referred to as *spurious* or *apparent* state dependence (Heckman and Willis, 1977; Heckman and Borjas, 1980; Heckman, 1981a; Heckman, 1981b; Honore and Kyriazidou, 2000).

For evidence of state dependence, therefore, an econometric model capable of separately identifying the causal effects of prior over-education and (observable and unobservable) individual characteristics is required. Following the standard approach in studies that test for state dependence (or the source of persistence) in labour market states (see, for example, Arulampalam et al. (2000), Stewart (2007), Oguzoglu (2011) and Mavromaras and McGuinness (2012)), a dynamic panel probit model is estimated. In particular, a dynamic random effects probit model is used to estimate the effect that being over-educated in the previous year has on an individual's current likelihood of being over-educated, with evidence it increases the likelihood considered indicative of state dependence in over-education, and, therefore, evidence that over-education leads to human capital depreciation. The results also identify the relative importance of individual heterogeneity (both observed and unobserved) and state dependence in the observed persistence in over-education. The model and empirical estimators used to derive estimates are outlined below.

6.4.1 Econometric model and estimators

For the dynamic random effects probit model, the underlying model is defined as:

$$Y_{it}^{*} = X_{it}^{'}\beta + \delta Y_{it-1} + \alpha_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 2,...,T \qquad [6.1]$$

where Y_{it}^{*} is the unobservable propensity to be over-educated for individual i at time t, X_{it} is a vector of controls for individual characteristics at t, β is a vector of coefficients associated with X_{it}, Y_{it-1} is an indicator for being over-educated in the previous year, δ is the estimated parameter of interest (used to test for state dependence), α_i is a time-invariant random variable representing unobserved effects for individual i, and ε_{it} are idiosyncratic errors that vary across i and t (i.e., the usual regression disturbances).[17] Since Y_{it}^{*} is unobserved in data, the binary variable examined is:

$$Y_{it} = \begin{cases} 1 \text{ if } Y_{it}^{*} > 0 \\ 0 \text{ otherwise} \end{cases} \qquad [6.2]$$

whereby an individual is observed over-educated when their propensity to be over-educated exceeds a certain threshold (here assumed to be zero).

As in the standard random effects probit model, it is assumed $\varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^{2})$ and ε_{it} is independent of X_{it} for all i and t. It is also typically assumed α_i is uncorrelated with X_{it} for all i and t; a critical assumption given any violation of this condition leads to biased and inconsistent parameter estimates (Arulampalam et al., 2000). For the study of over-education—where α_i is likely to contain factors associated with individuals' human capital (e.g., innate abilities) and X_{it} includes controls for educational attainment—this no correlation assumption is unsustainable. Thus, following the standard practice in empirical studies that estimate dynamic random effects probit models (see, for example, the studies listed above and Lee and Oguzoglu (2007), Clark and Kannellopoulos (2009), Arulampalam and Stewart (2009), Buddelmeyer, Lee and Wooden (2010), Jeon, Kalb and Vu (2011)), the Mundlak (1978) and Chamberlain (1984) approach is used to relax the no correlation assumption. That is, to allow for such correlation, the relationship between α_i and X_{it} is specified parametrically and incorporated into the model. This relationship is assumed to be as follows:

$$\alpha_i = \overline{X}_i^{'} a + u_i \qquad [6.3]$$

where $\overline{X}_i$ is a vector of means for the time-varying covariates for individual i over the entire period examined, and it is assumed $u_i \sim N(0, \sigma_u^{2})$ and u_i is independent of X_{it} and ε_{it} for all i and t (as such,

[17] The model assumes the coefficients β and δ are constant over time (i.e., constant over the eight-year period examined). Also, Y_{it-1} is actual over-education status in the previous year, not individuals' propensity to be over-educated.

u_i can be regarded as the residual time-invariant unobserved heterogeneity).[18, 19] Incorporating [6.3] into the model in [6.1] produces:

$$Y_{it}^{*} = X_{it}^{'}\beta + \delta Y_{it-1} + \bar{X}_{i}^{'}a + u_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 2,...,T. \quad [6.4]$$

If u_i is uncorrelated with the explanatory variables in the model, then it can be absorbed into the composite error term $\nu_{it} = u_i + \varepsilon_{it}$ and [6.4] can be estimated (as a standard random effects probit model) to produce unbiased and consistent parameter estimates. But, since the propensity to be over-educated in each period (Y_{it}^{*}) is correlated with time-invariant unobserved heterogeneity (u_i), the u_i will be correlated with over-education status in the previous year ($Y_{it\text{-}1}$). In estimating [6.4], $Y_{it\text{-}1}$ is therefore endogenous (i.e., correlated with the composite errors ν_{it}). Ultimately, the dynamic nature of the model means individuals' current over-education status depends on their experiences dating back to their initial over-education status. Thus, as stated by Heckman (1981c), estimating the parameters of this stochastic process with dependence in the outcome variables requires the process be initialised somehow. This is commonly referred to as the *initial conditions problem* (Heckman, 1981c; Arulampalam et al., 2000; Stewart, 2007).

Given its endogeneity, the previous over-education status of each individual needs to be explicitly modelled. And this is actually done in the model in [6.4], but just not for each individual's first period observed in the data (i.e., Y_{i1}). The initial conditions problem, therefore, reduces to the need to account for the endogeneity of each individual's initial over-education status in the data. If the problem is ignored (and Y_{i1} is implicitly assumed to be exogenous), then the result would be biased and inconsistent parameter estimates—in particular, an upwardly biased estimate of the state dependence in over-education (δ) (Heckman, 1981c; Stewart, 2007; Oguzoglu, 2011).[20, 21] In empirical

---

[18] In [6.3], a contains an intercept term and the coefficients corresponding to time-invariant covariates are set equal to zero.

[19] This approach is a form of quasi-fixed effects estimation; the Mundlak-Chamberlain controls—means of the time-varying covariates in [6.1]—are included in the model to control for the unobserved effects that may be correlated with the controls for individuals' (time-varying) characteristics (Lee and Oguzoglu, 2007; Vu, 2010). To fully relax the no correlation assumption requires a fixed effects estimator (Oguzoglu, 2011). But, the dynamic fixed effects probit model is considered too unreliable and recent advances in dynamic non-linear fixed effects models cannot be applied to analyses (such as this chapter) that use detailed datasets (Heckman, 1981c; Honore and Kyriazidou, 2000; Oguzoglu, 2011).

[20] Assuming initial conditions are exogenous is only valid if the start of the stochastic process is observed for each individual or the disturbances for the stochastic process are serially independent (Heckman, 1981c). Neither condition is true here. First, since the data cover only an eight-year period and contain individuals of all ages and career stages, the initial over-education status is not observed for each individual. Second, the disturbances in [6.4] are not serially independent: the composite error term ν_{it} is correlated between any two (different) time periods for a given individual, with this correlation a constant and given by $\rho = corr(\nu_{it}, \nu_{is}) = \dfrac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$ for $t, s = 2,...,T; t \neq s$.

[21] A similarly simple approach is to assume the stochastic process is in an equilibrium or steady state at the start of the period examined (Heckman, 1981c; Akay, 2012). This assumption, however, is inappropriate if the model contains time-varying covariates. Since [6.4] contains such time-varying covariates, the assumption cannot be made here.

studies, the initial conditions problem has been addressed using three main methods: the estimators proposed by Heckman (1981c), Orme (1997) and Wooldridge (2005).

Heckman estimator

The Heckman estimator addresses the initial conditions problem by specifying the following reduced form equation for the initial period:

$$Y_{i1}^{*} = \delta' Z_{i1} + \theta u_i + \varepsilon_{i1} \qquad [6.5]$$

where Z_{i1} is a vector of exogenous instruments consisting of X_{i1}, $\overline{X}_i$ (once again included to allow for correlation between time-invariant unobserved heterogeneity and the time-varying covariates) and variables capturing pre-sample information on time-varying exogenous factors that affect individuals' likelihood of being over-educated (i.e., measures for characteristics and experiences in the time before individuals are first observed in the HILDA Survey data). Together [6.4] and [6.5] specify a complete model for the dynamic over-education process; full information maximum likelihood estimation then produces unbiased and consistent parameter estimates.[22]

Initial conditions are identified in two ways. First, since the coefficients for X_{i1} in [6.5] are allowed to differ from those in [6.4], these covariates are able to identify their direct effects on the initial conditions and also act as instruments for the influence of previous over-education status. Second, initial conditions are identified by the variables with pre-sample information as these instrument directly for each individual's over-education history prior to Y_{i1}. As a result, the inclusion and choice of such pre-sample variables is critical for identifying the initial conditions. Hypothesis tests on the individual and joint statistical significance of the estimated coefficients for these pre-sample variables provide a means for assessing whether the instruments are indeed identifying the initial conditions. Further, the statistical significance of θ assesses whether the initial conditions are indeed endogenous in the model (where the null hypothesis is exogeneity).

In empirical studies, use of the Heckman estimator has been limited because the detailed programming required is not available as a standard program in software packages and because the computationally demanding procedure typically takes many hours (or days) to achieve convergence and produce results (Stewart, 2007; Arulampalam and Stewart, 2009; Clark and Kanellopoulos, 2009). It is for these reasons that alternative estimators were developed.

Orme estimator

The Orme estimator is a two-step procedure for addressing the initial conditions problem, similar to the standard sample selection correction method (as developed by Heckman (1979)). Essentially, it

[22] For further details on the Heckman estimator see Heckman (1981c), Arulampalam et al. (2000), Stewart (2007) and Arulampalam and Stewart (2009).

accounts for the correlation between the initial conditions (Y_{i1}) and time-invariant unobserved heterogeneity (u_i) by adding a correction term to the model in [6.4]. The two-step procedure is as follows. First, a probit model is used to estimate the following reduced form model for each individual's initial over-education status:

$$Y_{i1}^{*} = \delta' Z_{i1} + \eta_i \qquad [6.6]$$

where $\eta_i = \theta u_i + \varepsilon_{i1}$ (such that [6.6] is similar to [6.5] used by the Heckman estimator) and Z_{i1} is, once again, a vector of exogenous instruments consisting of X_{i1}, $\overline{X}_i$ and the pre-sample variables. Second, results from the estimation of [6.6] are used to derive an approximation for time-invariant unobserved heterogeneity that is uncorrelated with the initial conditions; this approximation is denoted e_i and defined as the probit generalised error:

$$e_i = E\left(\eta_i \mid Y_{i1}\right) = \frac{\left(2Y_{i1} - 1\right)\phi\left(\delta' Z_{i1}\right)}{\Phi\left[\left(2Y_{i1} - 1\right)\delta' Z_{i1}\right]} \qquad [6.7]$$

where $\phi(.)$ and $\Phi(.)$ denote the probability density and cumulative distribution functions of the standard Normal distribution. e_i is substituted for u_i in [6.4], such that the model becomes:

$$Y_{it}^{*} = X_{it}'\beta + \delta Y_{it-1} + \overline{X}_i'a + e_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 2,...,T. \qquad [6.8]$$

The standard random effects probit model is then used to estimate [6.8], and this produces unbiased and consistent parameter estimates.[23] Similar to the test of θ in the Heckman estimator, the statistical significance of the coefficient for e_i assesses whether the initial conditions are endogenous in the model (i.e., whether Y_{i1} is correlated with u_i).

Since it involves the estimation of standard probit and random effects probit models, the Orme estimator is more straightforward to implement than the Heckman estimator. And its estimation takes a fraction of the time. Monte Carlo evidence from Arulampalam and Stewart (2009) also indicates that results from the Orme estimator are definitely no worse, and in some cases actually better, than those derived using the Heckman estimator. For these reasons, the Orme estimator is considered a valid alternative for addressing the initial conditions problem.

Wooldridge estimator

Unlike the Heckman estimator (which models the density of (Y_{i1}, Y_{i2}, …, Y_{iT}) given X_i), the Wooldridge estimator models the density of (Y_{i2}, Y_{i3}, …, Y_{iT}) conditional on (Y_{i1}, X_i). Hence the key difference is the Wooldridge estimator specifies a model for time-invariant unobserved heterogeneity (α_i) given Y_{i1} and X_i, rather than specifying a model for the initial conditions (Y_{i1}) given X_i and α_i.

[23] For further details on the Orme estimator see Orme (1997), Arulampalam et al. (2000), Arulampalam and Stewart (2009) and Clark and Kanellopoulos (2009).

The Wooldridge estimator begins with the model in [6.1], and combines it with the following model for time-invariant unobserved heterogeneity:

$$\alpha_i = \varsigma Y_{i1} + Z_i^{'} a + \xi_i \qquad [6.9]$$

where (similar to the Mundlak-Chamberlain approach for relaxing the no correlation assumption) Z_i contains the vector of means $\overline{X}_i$, and it is assumed $\xi_i \sim N\left(0, \sigma_\xi^2\right)$ and ξ_i is independent of Y_{i1}. This results in the following model:

$$Y_{it}^{*} = X_{it}^{'}\beta + \delta Y_{it-1} + \varsigma Y_{i1} + \overline{X}_i^{'} a + \nu_{it}, \qquad i = 1,...,N; t = 2,...,T \qquad [6.10]$$

where $\nu_{it} = \xi_i + \varepsilon_{it}$. The standard random effects probit model is then used to estimate [6.10], and this produces unbiased and consistent parameter estimates.[24] Also, the statistical significance of ς assesses whether the initial conditions are endogenous in the model.

Of the three main methods used to address the initial conditions problem, the Wooldridge estimator is clearly the most straightforward to implement. And, similar to the Orme estimator, its estimation time is a fraction of that of the Heckman estimator without there being any adverse effects on the accuracy of its estimates (Arulampalam and Stewart, 2009). Not surprisingly, the Wooldridge estimator is widely used in empirical studies.

6.4.2 Empirical results

In this analysis, the Heckman, Orme and Wooldridge estimators are all used to estimate the above model for the probability of being over-educated. The dynamic structure of the model leads to some restrictions on the sample that can be examined. In particular, since estimation requires consecutive observations for individuals (to identify the effect of being over-educated in the previous year), there are two sample restrictions: first, individuals with only a single observation in the data are omitted from the analyses; second, for individuals with a break between observations (because they are a non-respondent or not in-sample in a particular year), only the first spell of consecutive observations is included in the analyses (to ensure their initial over-education status is appropriately modelled).[25] Ultimately, unbalanced panels are again examined, with the robustness of the results considered by replicating the analyses using balanced panels.[26] Also, similar to analyses in previous chapters, the model is estimated separately for males and females.

---

[24] For further details on Wooldridge estimator see Wooldridge (2005), Stewart (2007) and Arulampalam and Stewart (2009).
[25] To be clear, 'in-sample' refers to individuals who are employed, between 15 and 64 years of age, not a full-time student and not self-employed. Individuals are also treated as having a break between observations if, in a particular year, they have missing information for any of the explanatory variables used in the model (as they would consequently drop out of the analysis). Similar sample restrictions are applied in Arulampalam et al. (2000).
[26] These balanced panels require individuals to be a survey respondent, in-sample and have non-missing values for all the variables used in the model in each year.

With regards to the specification of the model, it is once again possible to control for a detailed set of individual characteristics. But, in using the Heckman estimator, the choice of explanatory variables—particularly, the number of covariates included—acutely affects the time taken for the estimator to achieve convergence and produce results. In response, the model controls for only the key characteristics thought to affect individuals' likelihood of being over-educated.[27] Specifically, the model contains measures for individuals' stock of human capital (*highest education level*, *labour market experience*, *tenure in current occupation*, *years unemployed*, *English proficiency* and *existence of a long-term health condition*), demographic characteristics (*age*, *ethnicity*, *marital status*, *number of children* and *remoteness of area of residence*) and year dummies.[28, 29] Also, as previously discussed, the model contains the Mundlak-Chamberlain controls: measures for the means of the time-varying covariates in the model, which are defined using all observations in the data for each individual (not just the observations for which individuals are in-sample).[30] Since the model is designed to control for unobserved individual heterogeneity, this choice of explanatory variables should not significantly affect estimates of the state dependence in over-education. Nevertheless, a model with an extended set of explanatory variables is estimated (using the Orme and Wooldridge estimators) to ensure the robustness of the results.[31]

For the Heckman and Orme estimators, the vector of exogenous instruments used to model initial over-education status (i.e., Z_{i1} in [6.5] and [6.6]) must also be defined. Recall, Z_{i1} must contain variables capturing pre-sample information on time-varying factors that affect Y_{i1}^{*}, where the choice of such variables is critical for identifying the initial conditions. In this analysis, Z_{i1} contains the following pre-sample information: *number of jobs last year*; *proportion time employed last year*; *proportion time unemployed last year*; and, *number of homes lived in during past ten years*. Each measure is time-varying and refers to the period prior to individuals being observed in the data: the first three measures refer to the (Australian financial) year immediately prior, while the last refers to the previous ten years. Further, each measure is likely to affect individuals' likelihood of being over-educated in the initial period: the first three may affect their stock of human capital and opportunities faced in labour markets, while the last is a proxy measure for their relative mobility or willingness to move for employment. Some measures for time-invariant family background characteristics are also included as such factors may indirectly affect individuals' human capital; these are: *employment status of father/mother*

[27] Given the following (restricted) specification, the Heckman estimator still takes approximately 24 hours to converge and produce results.
[28] Squared-terms for experience and occupation tenure (and a cubed-term for experience) capture the typically non-linear effects of experience and tenure.
[29] For further details on these explanatory variables see Appendix 5.1.
[30] Analysis of the *within* variation in the covariates listed above indicates that all except *ethnicity* exhibit variation over time.
[31] The additional measures included in this extended specification are: *number of post-school qualifications*; *tenure with current employer* (including a squared-term); *rest of household income* (including a squared-term); and, *State/Territory of residence*.

when individual aged 14; *whether father unemployed for six months or more when individual grew up*; *number of siblings*; and, *whether eldest child*. Each measure, though time-invariant, captures pre-sample information that may affect individuals' likelihood of being over-educated and, therefore, may aid in identifying the initial conditions.[32] Finally, as stated in the previous sub-section, Z_{i1} also contains measures for initial values of each covariate in the model (X_{i1}) and means of the time-varying covariates ($\overline{X}_i$).

Given these specifications, Table 6.6 presents results from estimating the dynamic model for over-education using the Heckman, Orme and Wooldridge estimators. Results from a pooled (dynamic) probit model are also presented; these ignore the potential effects of unobserved individual heterogeneity and endogenous initial conditions. Thus, they provide a means for examining how estimates of the state dependence in over-education are affected by accounting for such factors.[33] But, to enable comparisons, the estimated coefficients from the Heckman, Orme and Wooldridge estimators must be re-scaled; the original coefficients are multiplied by an estimate of $\sqrt{1-\rho}$, where ρ is the constant cross-period correlation in individuals' errors (as defined in a footnote above).[34] Tables A6.3.1 and A6.3.2 in Appendix 6.3 present the complete set of results for males and females, and Table A6.3.3 contains results of the initial conditions models that are estimated as part of the Heckman and Orme estimators.

Prior to examining the estimates of the state dependence in over-education, results from the diagnostic tests performed by each estimator are considered. First, the Heckman, Orme and Wooldridge estimators all provide an estimate of ρ; this measures the proportion of the unexplained variation in over-education status that can be attributed to unobserved individual heterogeneity (or factors absent from the model), and its statistical significance assesses the need to account for such unobserved heterogeneity in the model. If ρ is not statistically significant, then unobserved heterogeneity does not affect the parameter estimates and the pooled probit estimates are sufficient. For each of the estimators, the estimate of ρ is statistically significant (see results in Table 6.6). Results from the Orme and Wooldridge estimators indicate roughly 47 to 50 per cent of the unexplained variation in over-education status can be attributed to the unobserved heterogeneity, while results from the Heckman estimator are considerably higher and around 73 per cent for males

---

[32] Similar family background measures have been used as instruments in previous empirical studies that estimate dynamic random effects probit models (see, for example, Arulampalam et al. (2000), Lee and Oguzoglu (2007), Stewart (2007), Arulampalam and Stewart (2009) and Vu (2010)).

[33] Pooled probit model estimates are also consistent regardless of whether the assumption that X_{it} is exogenous (as required in the dynamic random effects probit model) holds; thus, they provide reasonably robust estimates of the state dependence in over-education that account for observed individual heterogeneity (Wooldridge, 2003; Lee and Oguzoglu, 2007).

[34] As outlined by Arulampalam (1999) and Stewart (2007), this re-scaling is necessary because the models use different normalisations to estimate parameters: pooled probit estimates are normalised using variance of the composite error term (i.e., $\sigma_v^2 = 1$), while random effects probit estimates are normalised using variance of the random error term (i.e., $\sigma_\varepsilon^2 = 1$).

and 64 per cent for females. This evidence clearly indicates the importance of controlling for unobserved individual heterogeneity in the dynamic model for over-education.

Table 6.6: Dynamic random effects probit model estimates for probability over-educated

Dependent variable: Over-educated $_t$	Males				Females			
	Pooled probit	Heckman	Orme	Wooldridge	Pooled probit	Heckman	Orme	Wooldridge
	Coeff.	*Re-scaled coeff.*	*Re-scaled coeff.*	*Re-scaled coeff.*	*Coeff.*	*Re-scaled coeff.*	*Re-scaled coeff.*	*Re-scaled coeff.*
Over-educated $_{t-1}$	2.848**	1.055**	1.400**	1.394**	2.895**	1.323**	1.504**	1.441**
	(0.042)	(0.096)	(0.087)	(0.088)	(0.043)	(0.124)	(0.092)	(0.093)
$\hat{e}_i$			0.715**				0.704**	
			(0.103)				(0.115)	
Over-educ. $_{initial}$				1.223**				1.218**
				(0.179)				(0.201)
Rho (ρ)		0.731**	0.471**	0.476**		0.641**	0.475**	0.499**
		(0.042)	(0.045)	(0.045)		(0.062)	(0.048)	(0.047)
Theta (θ)		1.220**				1.424**		
		(0.160)				(0.212)		
N	13,002	16,314	12,904	13,002	11,670	14,931	11,608	11,670
(No. individuals)	(3,183)	(3,226)	(3,160)	(3,183)	(3,077)	(3,121)	(3,057)	(3,077)
Log likelihood	-2,317.2	-2,151.9	-2,219.1	-2,243.0	-2,171.6	-1,964.1	-2,086.0	-2,111.5

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels; significance tests for the Heckman, Orme and Wooldridge estimators are based on the original (rather than re-scaled) coefficients.
Standard errors are reported in parentheses.
Re-scaled coefficients are the product of the original coefficients and estimate of $\sqrt{(1-\rho)}$ (using estimate of ρ derived from the model); original coefficients are not presented, but are available from the author on request.
All models contain the following controls: *human capital measures* (highest education, experience, occupation tenure, years unemployed, English proficiency, long-term health condition); *demographic characteristics* (age, ethnicity, marital status, number of children, remoteness of area); *year dummies*; and, *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).
See Tables A6.3.1 and A6.3.2 in Appendix 6.3 for the complete set of results for males and females; also, see Table A6.3.3 in Appendix 6.3 for results of the initial conditions models (i.e., initial over-education status) that are estimated as part of the Heckman and Orme estimators.

The second diagnostic test concerns the endogeneity of the initial conditions (i.e., each individual's initial over-education status). Recall, each estimator has a different test for this endogeneity: the Heckman estimator assesses the estimates of θ; the Orme estimator assesses the coefficient estimates for $\hat{e}_i$; and, the Wooldridge estimator assesses the coefficient estimates for Y_{i1}. The null hypothesis in each case is exogeneity; thus, statistically significant estimates are evidence of endogenous initial conditions. Once again the results from all three estimators concur, and these results indicate that the initial conditions are indeed endogenous in the model—that is, initial over-education status is correlated with unobserved individual heterogeneity. Failure to account for this endogeneity in the dynamic model for over-education would, therefore, result in biased and inconsistent parameter estimates. And, as a result, the Heckman, Orme and Wooldridge estimates of

the state dependence in over-education are expected to be significantly different from those based on the pooled probit model.

The final diagnostic test concerns the statistical significance of the pre-sample variables (or instruments) in Z_{i1} that the Heckman and Orme estimators use to identify the initial conditions. Results from the initial conditions models—presented in Table A6.3.3 in Appendix 6.3—indicate few of the pre-sample variables have statistically significant coefficients for the Heckman estimator, though they are jointly statistically significant at the 5 per cent level. For the Orme estimator, however, results are considerably improved as several variables have statistically significant coefficients and they are jointly statistically significant at the 1 per cent level. In addition, among the other explanatory variables in Z_{i1} (i.e., X_{i1} and $\overline{X}_i$), there are many more with statistically significant coefficients in the initial conditions models for the Orme estimator than for the Heckman estimator. Therefore, while the evidence indicates the instruments are indeed identifying the initial conditions for both estimators, it appears the Orme estimator is more accurately modelling these initial conditions. For this reason, the Orme estimates of the state dependence in over-education are preferred to those of the Heckman estimator.

With regards to the state dependence in over-education, the estimates in Table 6.6 all indicate that being over-educated in the previous year has a positive, statistically significant effect on an individual's current likelihood of being over-educated. The pooled probit estimates are clearly the largest, but this is not surprising given they do not account for the effects of unobserved individual heterogeneity and incorrectly assume the initial conditions are exogenous. These two factors therefore appear to explain a lot of the observed persistence in over-education. Among the more sophisticated estimators, the Orme and Wooldridge estimates are similar and the Heckman estimates are somewhat smaler, particularly for males. Since the Orme estimates are preferred to the Heckman estimates, it is reasonable to conclude that the Orme and Wooldridge estimates represent the most accurate estimates of the state dependence in over-education.[35] The robustness of these Orme and Wooldridge estimates is examined by replicating the analyses using balanced panels (results presented in Table A6.3.4 in Appendix 6.3) and using a model that controls for an extended set of individual characteristics (results presented in Table A6.3.5 in Appendix 6.3). Estimates of the state dependence in over-education are somewhat increased by using balanced panels, and they become relatively similar in size for males and females. Meanwhile, use of an extended specification leads to estimates that are virtually unchanged. The preferred estimates of state dependence in over-education, therefore, appear quite robust.

[35] Evidence from Monte Carlo simulations in Arulampalam and Stewart (2009)—evidence that, as the initial conditions problem worsens, the relative bias associated with the Heckman estimates of state dependence increases more quickly than that of the Orme and Wooldridge estimates—provides further support for this preference.

Since the non-linearity of probit models means the coefficient estimates in Table 6.6 cannot be used to assess the magnitude of these state dependence estimates, average partial effects are calculated. The average partial effects are essentially the average of the individual marginal effects. That is, given the estimated coefficients and individuals' actual values for the explanatory variables in the model, two probabilities are calculated for each individual: the probability of being over-educated when the indicator for being over-educated in the previous year is set equal to zero (denoted P_0); and, the probability of being over-educated when the indicator for being over-educated in the previous year is set equal to one (denoted P_1). The difference between these two probabilities is then averaged across all individuals to produce the average partial effects.[36] Table 6.7 presents the probabilities and average partial effects based on the estimates from the pooled probit and dynamic random effects probit models, along with the transition rates and elementary estimates of state dependence previously presented (in Table 6.1).[37]

Table 6.7: Estimates of state dependence in over-education

	Males			Females		
	(I) Raw data	(II) Pooled probit	(III) Dynamic RE probit (Wooldridge)	(I) Raw data	(II) Pooled probit	(III) Dynamic RE probit (Wooldridge)
$Pr(O_t=1 \mid O_{t-1}=0)$ [P_0]	0.027	0.033	0.061	0.028	0.036	0.080
$Pr(O_t=1 \mid O_{t-1}=1)$ [P_1]	0.845	0.802	0.370	0.865	0.821	0.428
State dependence [$P_1 - P_0$]	0.818	0.769	0.309	0.837	0.785	0.348
Ratio [P_1 / P_0]	31.3	24.3	6.1	30.9	22.8	5.4
N	17,359	13,002	13,002	16,637	11,670	11,670
(No. individuals)	(4,260)	(3,183)	(3,183)	(4,240)	(3,077)	(3,077)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures in column (I) are based on the raw data (as previously reported in Table 6.1), while figures in columns (II) and (III) are derived from results of the dynamic models for over-education.
Estimates differ across the columns for the following reasons: estimates in (I) do not account for any individual heterogeneity; estimates in (II) control for observed individual heterogeneity; and, estimates in (III) control for observed and unobserved individual heterogeneity (and the endogeneity of each individual's initial over-education status). Hence, column (III) presents the most reliable estimates of state dependence in over-education.

As previously discussed, the observed persistence of over-education in the raw data suggests substantial state dependence in over-education: the elementary estimates (in column (I) of Table 6.7) indicate that being over-educated in the previous year increases the likelihood an individual is currently over-educated by roughly 82 percentage points for males and 84 percentage points for

[36] These average partial effects are arguably more interpretable than simply calculating the marginal effects at the mean values of the explanatory variables (as sometimes done in empirical studies) because the use of mean values may not correspond to real individuals (especially since the model contains many categorical variables) (Lee and Oguzoglu, 2007).
[37] The average partial effects for the dynamic random effects probit models are based on the Wooldridge estimates; these are used (instead of the equally preferred Orme estimates) because they were generated using larger samples of individuals. Not surprisingly, average partial effects based on the Orme estimates are essentially identical, while those based on the Heckman estimates are somewhat reduced (these results are not presented, but are available from the author on request). Average partial effects based on results from the robustness analyses are also calculated and presented in Table A6.3.6 in Appendix 6.3; the results are qualitatively very similar to those in Table 6.7.

females (or, compared to those not previously over-educated, it makes them roughly 31 times more likely to be currently over-educated). Such large effects, however, do not account for (observed and unobserved) individual heterogeneity that may explain why individuals are persistently over-educated. Using pooled probit models to control for observed heterogeneity, the estimates are only marginally reduced: previous over-education now increases the likelihood an individual is currently over-educated by around 77 percentage points for males and 79 percentage points for females. Using dynamic random effects probit models to then also control for unobserved heterogeneity (and the endogeneity of each individual's initial over-education status), leads to significantly reduced estimates. But, despite the reduction, these estimates of the (true) state dependence in over-education remain significant: being over-educated in the previous year increases the likelihood an individual is currently over-educated by roughly 31 percentage points for males and 35 percentage points for females (or it makes them roughly six times more likely to be currently over-educated). With regards to the source of the observed persistence in over-education, the relative importance of these factors is as follows: observed individual heterogeneity accounts for around 6 per cent of the observed persistence; unobserved individual heterogeneity for 54 per cent; and, (true) state dependence for 40 per cent.[38] Thus, for approximately 60 per cent of the individuals observed persistently over-educated it is certain enduring characteristics that cause this persistence. For the remaining 40 per cent, the cause is the previous experiencing of over-education (i.e., state dependence).

The empirical analyses in this section, therefore, have found evidence of state dependence in over-education. And this is considered evidence of a link between over-education and human capital depreciation. That is, for some individuals, it appears time spent over-educated leads to depreciation of their (under-utilised) human capital.

6.5 Prior over-education and wages of well-matched individuals

To further consider whether over-education leads to human capital depreciation, this section examines the relationship between prior over-education and the wages of currently well-matched individuals. The effect that being over-educated in the previous year has on their current wages is estimated, with evidence that it reduces wages considered indicative of individuals having reduced labour productivity due to human capital depreciation. The analysis focuses on the effect of prior over-education among individuals who recently exited over-education (i.e., moved from over-educated to well-matched between adjacent waves of the data) because it is, arguably, during this

[38] These figures are calculated using the results in Table 6.7; they are the differences between the respective estimates of state dependence divided by the estimates in column (I). For example, the relative importance of (true) state dependence is calculated as the difference between the estimates in columns (I) and (III) divided by the estimates in column (I).

short-term period that the effects of human capital depreciation will be most pronounced. Thus, it is the best framework for examining whether any human capital depreciation occurs.[39]

6.5.1 Econometric model and estimators

For results to be interpreted as evidence of human capital depreciation, it is important, once again, that the estimates represent causal effects. The DID matching estimator (as defined in Section 5.3.3) is therefore used to derive the estimates. It is used because the assumptions necessary for it to derive valid causal effects estimates are considerably weaker than those required when using regression analyses (i.e., compared to the fixed effects estimator) and because it produced the wage penalty estimates most likely to represent causal effects in Chapter 5. Use of the DID matching estimator is also possible because, based on the sample examined (individuals well-matched at *t*), the analysis involves matching a small *treatment group* (those over-educated at *t–1*) to a large *control group* (those well-matched at *t–1*) and such comparisons are sufficient for the identification of the ATT (the causal effect of prior over-education among the currently well-matched).[40]

6.5.2 Empirical results

As outlined above, this section examines the sample of individuals well-matched in a given year (denoted *t*) and is concerned with the effect of their over-education status in the previous year (*t–1*). Two observations are therefore required for each individual, and so two-year panels of data are constructed using the adjacent waves of the HILDA Survey data (with these panels then pooled together for the analysis). This results in a sample of 23,087 individuals well-matched at *t*, where 663 are over-educated at *t–1* and 22,424 are well-matched at *t–1*.[41] For the preliminary analysis of wages, which considers average wages by over-education status at *t–1* and *t–2*, three-year panels are also constructed (for individuals well-matched at *t* and *t–1*). Not surprisingly, these contain considerably smaller samples of individuals. Consistent with previous analyses in this study, real hourly wages (expressed in 2001-dollars) are examined and effects are estimated separately for males and females. In this analysis, it is also necessary to control for any effects associated with occupation change because all individuals who recently exited over-education did so via a change in occupation (see results in Table A6.1.2 in Appendix 6.1) and such changes in occupation may, on their own, have significant effects on individuals' wages. For example, voluntary changes in occupation may be

[39] Descriptive statistics (and pooled OLS estimates) are also considered for individuals over-educated two years ago (i.e., the effect of being over-educated at *t–2* is examined for individuals well-matched at *t* and *t–1*). But, apart from such preliminary analyses, this analysis does not examine the dynamics of the effect of prior over-education (i.e., whether the effect dissipates with greater time spent well-matched) and, therefore, does not consider whether human capital depreciation associated with over-education is permanent; this is a suitable avenue for future research.

[40] The size of the sample examined (and these *treatment* and *control* groups) is defined in the following sub-section.

[41] Individuals may have multiple observations in this sample; thus, the data examined are essentially unbalanced panels.

associated with wage increases (i.e., a positive effect) for individuals. If not controlled for, such effects could lead to biased estimates of the effect of prior over-education. DID matching estimates, therefore, are also derived using the sub-sample who changed occupation between *t–1* and *t*: individuals who changed occupation and moved from over-educated to well-matched are compared to those who changed occupation and were well-matched in both periods.

Prior to examining the DID matching estimates, Table 6.8 presents means and standard deviations for the hourly wages of currently well-matched individuals by prior over-education status, along with the differences in their mean wages. These differences are elementary estimates of the effect of prior over-education, with panel A using the two-year panels to estimate the effect of being over-educated in the previous year and panel B using the three-year panels to estimate the effect of being over-educated two years ago. All differences are statistically significant at the 1 per cent level, except that for females in panel B. The figures in panel A indicate individuals over-educated in the previous year earn less than those who were well-matched; the difference is, on average, 13 per cent for males and 11 per cent for females. Meanwhile, results in panel B indicate such wage differences may persist over time, as individuals over-educated two years ago continue to earn less (than those well-matched in all three years): the wage difference is relatively unchanged for males at approximately 14 per cent, but somewhat reduced for females at roughly 7 per cent. The preliminary evidence, therefore, indicates that among currently well-matched individuals there is a wage penalty associated with (recent) prior over-education.[42]

[42] Similar evidence also arises from the estimation of an ORU earnings function akin to the model in [5.3], where the model contains the same extensive set of controls for individual characteristics except O_{it} is replaced with O_{it-1} (or O_{it-2}) and the model is estimated only for individuals well-matched at *t* (and, to control for any effects of occupation change, it is then also estimated using the sample who changed occupation between the two years). Pooled OLS estimates, which are presented in Table A6.4.1 in Appendix 6.4, resemble the estimates in Table 6.8: being over-educated at *t–1* leads to a wage penalty of roughly 9 to 11 per cent for males and 7 to 8 per cent for females, while being over-educated at *t–2* results in a wage penalty of 13 to 15 per cent for males and 4 to 7 per cent for females. And, importantly, these wage penalty estimates remain statistically significant after controlling for the effects of occupation change. For estimates more likely to represent causal effects, the model is also estimated using the fixed effects estimator (though the model with O_{it-2} cannot be estimated because the three-year panels contain very few individuals with multiple observations). The fixed effects estimates, which are presented in Table A6.4.2 in Appendix 6.4, are considerably smaller than the previous estimates: the wage penalty for being over-educated at *t–1* is 1 to 2 per cent (and not statistically significant) for males and 4 to 5 per cent for females. Important limitations of these fixed effects estimates, however, are that the two-year panels contain few individuals with multiple observations (which adversely affects the ability to interpret the estimates as causal effects) and that the estimates cannot control for effects of occupation change (because restricting the sample to those who changed occupation between the two years results in very few individuals with multiple observations in the data). For these reasons, the fixed effects estimates are not considered to be reliable evidence on the causal effect prior over-education has on the wages of currently well-matched individuals. Of course, due to the bias that results from unobserved individual heterogeneity, the same also applies to the pooled OLS estimates (and the elementary estimates in Table 6.8).

Table 6.8: Average wages by prior over-education status and gender—Well-matched individuals

	Males				Females			
	Mean	*Difference*		*N*	Mean	*Difference*		*N*
		($)	(%)			($)	(%)	
A. Sample: Individuals well-matched at *t* (two-year panels)								
Well-matched $_{t-1}$	23.42			*11,506*	20.26			*10,918*
	(12.66)				(13.57)			
Over-educated $_{t-1}$	20.37	-3.05	-13.0	*306*	18.02	-2.24	-11.1	*357*
	(10.70)				(7.84)			
N				*11,812*				*11,275*
B. Sample: Individuals well-matched at *t* and *t-1* (three-year panels)								
Well-matched $_{t-2}$	24.10			*8,074*	20.69			*7,572*
	(12.51)				(13.21)			
Over-educated $_{t-2}$	20.77	-3.33	-13.8	*171*	19.20	-1.49	-7.2	*189*
	(9.21)				(7.50)			
N				*8,245*				*7,761*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures in panel A are based on pooled two-year panels of adjacent waves of the data; figures in panel B are based on pooled three-year panels of adjacent waves of the data.
Real hourly wages examined; see Table A5.1.1 in Appendix 5.1 for details on the derivation of real hourly wages.
Standard deviations reported in parentheses.
All differences are statistically significant at the 1% level, except that for females in panel B.

The DID matching estimator, as defined in [5.16], uses the two-year panels to examine the changes in individuals' wages between the two years.[43] Specifically, for the sample well-matched at *t*, individuals over-educated at *t–1* are matched to individuals well-matched at *t–1* (based on individual characteristics at *t–1*) and then the difference in their changes in wages (between *t–1* and *t*) are used to estimate the causal effect of being over-educated at *t–1*. This matching is similar to that performed in Chapter 5, whereby: quasi-exact matching is performed on individuals' highest education level and gender; logit models for the probability over-educated (at *t–1*) are estimated to derive propensity scores, and separate models are estimated for males and females; specifications of the logit models are determined using the Dehejia and Wahba (2002) algorithm that considers results from balancing tests; and, the caliper method is used during matching to enforce the CSA. For the DID matching to provide valid causal effects estimates the following identifying assumption, similar to that defined in [5.15], must hold: conditional on a set of observed characteristics (or propensity scores), the wages of individuals well-matched in both periods must have evolved over the years in the same way the wages of individuals over-educated at *t–1* would have evolved had they instead been well-matched at *t–1*. Consistent with the discussion above, this condition is most likely to hold when the effects associated with occupation change are controlled for.

Table 6.9 presents the DID matching estimates. Estimates are derived using nearest neighbour and kernel matching (i.e., both are used to determine the weights in [5.16]), and panel A

[43] The outcome actually examined is the change in the log of hourly wages because then, similar to the use of logs in regression analyses, the differences across time periods represent percentage changes in wages.

contains the estimates derived using all individuals well-matched at *t* and panel B contains those derived using the sub-sample who changed occupation between *t–1* and *t*. For each estimate, the logit models used to derive propensity scores contain extensive controls for individual characteristics—measures for human capital, recent labour market experiences, demographic characteristics and family background (see Notes to Table 5.3 for exact list of variables in these categories)—along with year dummies and interaction terms aimed at satisfying the balancing tests. Also, based on the estimated propensity scores, there appears to be sufficient overlap in the data between individuals over-educated at *t–1* and individuals well-matched at *t–1* to enable matching (i.e., the CSA appears satisfied), and then almost all the previously over-educated individuals are successfully matched (i.e., less than 1 per cent cannot be matched) and adequate numbers of well-matched individuals are used in the matching (i.e., estimates are not overly reliant on small numbers of well-matched individuals to identify the counterfactual wages). In addition, results from the balancing tests indicate no significant imbalance exists—no statistically significant differences in (mean) characteristics—in the matched samples used to derive the estimates.[44, 45]

Table 6.9: Wage effects of prior over-education among well-matched individuals—Difference-in-differences matching estimates

	Males		Females	
	DID matching (*Nearest neighbour*)	DID matching (*Kernel*)	DID matching (*Nearest neighbour*)	DID matching (*Kernel*)
A. All individuals well-matched at *t*				
Difference in (log) real hourly wage	-0.054	-0.048*	-0.057	-0.050*
	(0.038)	(0.023)	(0.035)	(0.025)
N	532	8,929	645	8,159
(No. over-educated $_{t-1}$)	(278)	(278)	(335)	(335)
B. Restricted sample: Individuals who changed occupation between *t-1* and *t*				
Difference in (log) real hourly wage	-0.011	-0.010	-0.023	-0.025
	(0.047)	(0.034)	(0.044)	(0.032)
N	440	930	536	947
(No. over-educated $_{t-1}$)	(278)	(278)	(335)	(335)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Bootstrap standard errors based on 1,000 replications reported in parentheses.

The estimates in panel A indicate that being over-educated in the previous year reduces currently well-matched individuals' wages by approximately 5 per cent (for both males and females).

[44] As in Chapter 5, three different balancing tests—the tests defined by Dehejia and Wahba (2002), Smith and Todd (2005) and Ho et al. (2007)—have been performed in this analysis. For further details on these balancing tests see Section 5.4.3.
[45] These results—the logit model results with the chosen specifications, the graphs establishing sufficient overlap exists in the propensity scores of the over-educated and well-matched individuals, the results indicating the estimates are not overly reliant on small numbers of well-matched individuals, and the results from the balancing tests—are not presented, but are available from the author on request.

Only the estimates derived using kernel matching are statistically significant—as the use of more information (or larger samples) to estimate the counterfactual wages results in them being more precisely estimated—but the nearest neighbour matching estimates are similar in magnitude and therefore, despite not being statistically significant, they are also regarded as evidence of prior over-education leading to a 5 per cent wage penalty. Estimates in panel B, however, are smaller in magnitude, with the wage penalty 1 per cent for males and 2 to 3 per cent for females, and are not statistically significant. Since the estimates in panel A are likely biased (as they do not account for the effects of occupation change) and the estimates in panel B are not significant, these results are interpreted as providing no evidence of (recent) prior over-education reducing the wages of currently well-matched individuals. This analysis, therefore, has found no empirical evidence of over-education leading to human capital depreciation.

The empirical test performed here, however, has an important limitation that may explain this finding. Specifically, the sample of individuals examined—those who recently exited over-education—is not representative of all over-educated individuals and such differences may then affect whether the test can actually derive evidence of human capital depreciation. For instance, if the individuals who recently exited are the more able or more highly productive among those over-educated, then they may be less susceptible to human capital depreciation while over-educated. Alternatively, they may be the individuals who have experienced the least human capital depreciation while over-educated; hence their ability to obtain a well-matched job. Results from supplementary analyses find this may indeed be the case. In particular, recall from Section 5.4.1, it was found that, compared to all over-educated individuals, the sample who recently exited over-education contains a higher proportion of younger individuals (aged 15 to 29 years), individuals with less labour market experience (and less occupation tenure and employer tenure), more highly educated individuals (i.e., those with a Bachelor Degree or higher) and more individuals with parents who had a post-school qualification.[46] These results (where the latter two differences may reflect individuals' ability levels) suggest the sample who recently exited may be less susceptible to human capital depreciation and have spent less time over-educated, which thereby means that within this sample there is less potential for human capital depreciation to have occurred. Hence, results from the empirical test performed in this section are likely to under-estimate the extent to which time spent over-educated leads to depreciation of individuals' (under-utilised) human capital.

[46] For details on these analyses and the complete list of differences in observable characteristics see Section 5.4.1.

6.6 Discussion and conclusion

The aim of this chapter was to consider whether instances of over-education are short-term labour market disequilibria that have no enduring effects. Empirical evidence was derived from two sets of analyses. The first considered whether over-education is short-term disequilibria: descriptive statistics on individuals' transitions to and from over-education and their durations over-educated were calculated, and then the over-education wage penalty was estimated by duration over-educated. Since it was again assumed that wages reflect the labour productivity achieved in the workplace, finding the wage penalty disappears with time spent over-educated was assumed to be evidence of instances of over-education being resolved by firms adjusting jobs. The second set of analyses considered whether over-education has enduring effects, and in particular, whether it leads to human capital depreciation. Two empirical tests were performed: the first tested for state dependence in over-education, whereby finding such state dependence was assumed to be evidence of human capital depreciation; the second estimated the effect that prior over-education has on the wages of currently well-matched individuals, whereby finding it reduces wages was assumed to be evidence of individuals having reduced labour productivity due to human capital depreciation.

For the eight-year period examined, it is not the case that individuals are only temporarily identified as over-educated and the over-education wage penalty has not been found to disappear with time spent over-educated. Instead, it was found that over-education is persistent for a sizeable group of individuals—generally, 30 to 40 per cent of the over-education in each year corresponds to individuals who are persistently over-educated, with the majority of these individuals over-educated for in excess of ten years—and the over-education wage penalty is relatively constant over time spent over-educated (which is evidence that those identified as persistently over-educated still have under-utilised human capital and, therefore, remain over-educated even after such extended periods). State dependence in over-education was also found: being over-educated in the previous year increases the likelihood an individual is currently over-educated by roughly 31 percentage points for males and 35 percentage points for females (or it makes them roughly six times more likely to be currently over-educated). Moreover, it was found that for approximately 40 per cent of the individuals observed persistently over-educated it is state dependence that causes this persistence. The final empirical test then found no evidence that prior over-education reduces the wages of currently well-matched individuals (though this may be due to a limitation of the test performed). Nevertheless, given the evidence of state dependence, it was found that over-education can lead to human capital depreciation. This chapter, therefore, has found that instances of over-education are not merely short-term labour market disequilibria that have no enduring effects.

The key implication from this finding is that over-education is not merely a by-product of adjustment processes in dynamic and well-functioning labour markets. Instead, it represents more serious labour market failures, and government policy interventions appear necessary to prevent and resolve such instances of over-education. Also, since it was found over-education can lead to human capital depreciation, the costs of over-education are greater than first thought as they not only arise while individuals are over-educated, but also in future employment where individuals may appear well-matched (i.e., their labour productivity will be lower compared to if they had not previously been over-educated). And this then increases the importance of government policy interventions to prevent over-education. In addition, since some individuals were found to be only temporarily over-educated, it appears government policy interventions cannot simply be targeted at all over-educated individuals collectively. Instead, differentiating between the individuals persistently over-educated and those likely to be only temporarily over-educated and then tailoring interventions specifically for each group appears necessary. Government policy interventions that resolve (and prevent) instances of persistent over-education appear likely to result in the greatest benefit for the economy.

In the over-education literature, there is little empirical evidence regarding the dynamics of over-education and, given the typically static analyses performed, it is unclear whether examining labour markets from a dynamic perspective diminishes the importance of over-education as labour market failures. This chapter, therefore, has made several contributions to the literature. Specifically, it has provided evidence that over-education is persistent for a sizeable group of individuals. It has provided the first evidence on the dynamics of the over-education wage penalty, both for individuals who remain over-educated over time and for those who exit to well-matched employment. It has also provided the first evidence of state dependence in over-education, which is considered evidence of a link between over-education and human capital depreciation. Ultimately, this chapter has provided empirical evidence that over-education is not merely a result of the inherently dynamic nature of labour markets. With a dynamic view of labour markets, therefore, instances of over-education continue to represent labour market failures.

The empirical tests performed in this chapter have some limitations. First, the descriptive analyses may not capture all of the variation in over-education status over time as data on individuals' changes in occupations, rather than jobs, are examined. This may lead to the persistence in over-education being overstated. Further evidence on duration over-educated is therefore needed. Second, and similar to the analyses in Chapter 5, *average* effects are estimated. This means, for example, the evidence that there is, on average, a statistically significant over-education wage penalty that remains relatively constant over time spent over-educated does not necessarily guarantee that all the over-educated individuals (in each duration over-educated category) experience such a wage penalty and, therefore, still have under-utilised human capital. Further examining the over-education wage penalty

by individuals' durations over-educated is a suitable avenue for future research. Similarly, the effects of persistent over-education—for instance, examining whether individuals' job satisfaction and life satisfaction levels vary by duration over-educated—warrant further consideration.

A further limitation is that the evidence of state dependence is not necessarily indicative of human capital depreciation. Such evidence may instead be the result of discrimination or stigma effects associated with over-education (e.g., firms mistakenly believing previously over-educated individuals are less productive, which adversely affects the job offers received by these individuals) or the fact individuals face costs of exiting over-education that exceed the benefits (e.g., the cost of geographical relocation exceeding the expected wage increase). Since the additional empirical test produced no evidence to support the link between over-education and human capital depreciation, further examination of whether such a link exists is a suitable avenue for future research. Evidence individuals' human capital depreciates while over-educated also affects the identification and interpretation of over-education, and so future research could also examine whether such human capital depreciation means some individuals are incorrectly identified as over-educated. Finally, given it was found that for roughly 60 per cent of the individuals observed persistently over-educated it is certain enduring characteristics that cause this persistence—that is, individual heterogeneity rather than state dependence—it appears possible that some of these individuals may remain over-educated because it is their preferred outcome. That is, as discussed in Sections 2.2 and 2.6, some individuals may trade wages for non-pecuniary benefits of employment or improved working conditions, and therefore could be deemed voluntarily over-educated. Such a possibility is considered in Chapter 7. The existence of such voluntary over-education then affects whether, as argued above, persistent over-education should be the target of government policy interventions. It is also possible that some of these individuals remain over-educated due to other enduring characteristics, such as innate ability level, quality of educational institution attended and qualification vintage, and so these factors should also be considered in future research.

Chapter 7

Voluntary over-education

7.1 Introduction

The preceding chapters have presented evidence of over-education in Australian labour markets and found it is not merely a by-product of adjustment processes in labour markets. Given the evidence of its persistence, a reasonable question is whether all the instances of over-education actually represent labour market failures. That is, since many attributes of jobs (e.g., work hours, job security and required effort), and not just the wage, can affect individuals' utility levels, it is possible some over-educated individuals have obtained jobs that maximise their (expected) utility levels and, therefore, achieved their preferred outcome. Such individuals, referred to here as *voluntarily over-educated,* would trade wages for non-pecuniary benefits of employment or improved working conditions and, as a result, accept jobs for which they are over-educated.[1] This chapter, therefore, considers the research question: *are some individuals voluntarily over-educated (i.e., trading wages for non-pecuniary benefits of employment or improved working conditions)?*

For relevant empirical evidence, it is first necessary to develop a method by which instances of voluntary over-education, if they exist, can be identified. This chapter proposes a distinction based on individuals' job satisfaction levels and their desire for a new job (or intentions to quit current job): it is assumed over-educated individuals who report being both highly satisfied with their job and highly unlikely to quit are *voluntarily over-educated*, while the remainder are *involuntarily over-educated*.[2] The method and resultant estimates must then be validated. If valid, there should be evidence of trade-offs between wages and other job attributes among the individuals identified as voluntarily over-educated, and hence the relationship between voluntary over-education and job attributes is examined. Specifically, the differences in job attributes between voluntarily over-educated and well-matched individuals, and the differences between the involuntarily over-educated and well-matched, are estimated. Finding that, compared to well-matched individuals, the voluntarily over-educated experience wage penalties and improvements in other job attributes and that the involuntarily over-educated incur wage penalties without such improvements is assumed to validate the empirical

[1] As stated in Chapter 2, such occurrences are consistent with *compensating wage differentials* in labour markets.
[2] The involuntarily over-educated are individuals who face constraints, such as family commitments, geographic location, poor health or labour market discrimination, which mean that, among their available (constrained) choices, accepting an over-educated job is the preferred outcome (i.e., the best job they can obtain, and it is preferred to non-employment).

identification of voluntary over-education. And, therefore, it would be evidence that some individuals are indeed voluntarily over-educated.[3]

A number of job attributes are considered, including: the wage; hours worked (and whether they are individuals' preferred hours); type of employment contract; job security; degree of flexibility to decide when to work (including whether the job has flexible start and finish times); degree of autonomy; and, degree of stressfulness. Since job attributes are treated as the dependent variables in the analyses and they vary between being measured as continuous, ordinal and binary, three different econometric models and estimators are used: the linear fixed effects, fixed effects ordered logit and fixed effects probit estimators.[4] Each uses a regression framework and the panel data to control for (observable and unobservable) individual heterogeneity, thereby ensuring the estimated differences in job attributes are not confounded by other factors (i.e., they reflect only the differences in job attributes that arise from being voluntarily (and involuntarily) over-educated rather than well-matched). Since the analyses produce evidence of voluntarily over-educated individuals, the chapter concludes by (briefly) considering the supplementary issue of whether such voluntary over-education explains the persistence in over-education (as found in Chapter 6). The correlation between persistent and voluntary over-education is examined, whereby the results identify the proportion of persistently over-educated individuals who are actually voluntarily over-educated.

This chapter proceeds as follows. Section 7.2 briefly recaps existing evidence relevant to the possibility of voluntary over-education. Section 7.3 discusses the empirical identification of voluntary over-education and presents estimates of its incidence. Section 7.4 assesses the validity of these estimates by examining the relationship between voluntary over-education and various job attributes. Section 7.5 examines the correlation between persistent and voluntary over-education. Section 7.6 concludes the chapter.

7.2 Existing evidence

Recall from Chapter 2, few studies have acknowledged the role individuals' preferences may play in the incidence of over-education and, as a result, there is little empirical evidence on the existence of

[3] This empirical test adopts a static approach to labour markets as it is assumed evidence of *contemporaneous* trade-offs—wage penalties and improvements in other job attributes that are experienced in the same time period—validates the existence of voluntary over-education. It represents a simplification because, in reality, individuals are likely to make employment decisions with respect to maximising their *lifetime* utility levels.

[4] A fixed effects probit estimator is preferred to the conditional fixed effects logit estimator because in using the latter there is a significant loss of information that adversely affects the efficiency of estimates and it is difficult to calculate partial effects estimates (Wooldridge, 2003; Cameron and Trivedi, 2005; Jones and Schurer, 2011). Nevertheless, the conditional fixed effects logit estimator is used in sensitivity analyses. Also, since it is difficult to derive consistent estimates for the fixed effects probit model (due to the incidental parameters problem) (Wooldridge, 2003; Greene, 2004; Cameron and Trivedi, 2005), they are instead approximated using a random effects probit estimator with Mundlak-Chamberlain controls. These issues are further discussed in Section 7.4.1.

voluntary over-education. Generally, evidence in the over-education literature—such as findings that, compared to the well-matched, over-educated individuals are less satisfied with their jobs and more likely to want to quit, be searching for a new job and experience a voluntary job separation—suggests over-education is involuntary for most individuals, but does not preclude the possibility it is voluntary for some. Two recent studies attempted to examine the issue more directly. Similar to this chapter, McGuinness and Sloane (2011) and Mavromaras et al. (2011) recognised that trade-offs between wages and other job attributes may mean some over-educated individuals are satisfied with their jobs and therefore voluntarily over-educated. Each study consequently examined the relationship over-education had with job satisfaction and job attributes. Ultimately, they found little evidence of voluntary over-education: while some of the over-educated had increased job security and job flexibility, over-education again tended to reduce job satisfaction levels. The results, however, are subject to some important limitations. In particular, both studies did not distinguish between voluntarily and involuntarily over-educated individuals prior to considering job attributes for evidence of trade-offs (i.e., estimates were derived for all over-educated individuals combined). Such an approach is unlikely to produce evidence of voluntary over-education because the estimates will tend to reflect the circumstances of involuntarily over-educated individuals (as voluntary and involuntary over-education have opposite effects and the existing evidence suggests over-education is involuntary for most individuals).[5]

This chapter addresses each of the limitations of McGuinness and Sloane (2011) and Mavromaras et al. (2011), and contributes in several ways to the over-education literature. Specifically, it is the first study to attempt to estimate the incidence of voluntary over-education. It is also the first to examine the relationship between such voluntary over-education and job attributes. And, in doing so, provides the first evidence on the particular job attributes for which voluntarily over-educated individuals trade wages (i.e., their reasons for accepting over-educated jobs). Finally, it is the first study to examine the correlation between persistent and voluntary over-education.

7.3 Estimating the incidence of voluntary over-education

In this chapter, individuals' job satisfaction levels and their desire for a new job (or intentions to quit current job) are used to empirically identify instances of voluntary over-education. This is done because, if the current job is indeed their preferred outcome, it is reasonable to expect voluntarily over-educated individuals would report being highly satisfied with the job and not wanting to leave it,

[5] Recall from Section 2.6, the other important limitations were: McGuinness and Sloane (2011) did not account for the effects of unobserved individual heterogeneity; Mavromaras et al. (2011) used the RM method to identify over-education (and used its changes over time to identify random effects probit estimates, which is inappropriate); and, both studies dichotomised the ordinal measures for job satisfaction and job attributes and then estimated probit models (rather than models designed for ordinal-valued measures). These limitations may lead to inefficient and biased estimates.

at least in the near future. On the other hand, involuntarily over-educated individuals would be unsatisfied with their current job and, most likely, seeking a new one. The chapter proposes, therefore, that over-educated individuals who report being both highly satisfied with their job and highly unlikely to quit (in the near future) can be deemed *voluntarily over-educated*, while the remainder are *involuntarily over-educated*.[6, 7]

Information in the HILDA Survey data on individuals' overall job satisfaction levels and intentions to quit their current job (in the next 12 months) is used to perform the distinction. The job satisfaction levels are measured on a scale from 0 (totally dissatisfied) to 10 (totally satisfied), while intentions to quit are measured as probabilities from 0 per cent (no chance) to 100 per cent (absolute certainty). Given these measures, thresholds must be defined for 'highly satisfied' and 'highly unlikely to quit'. Choosing the values for such thresholds is somewhat arbitrary—for example, is a job satisfaction level of 7 sufficiently high for an individual to be classified 'highly satisfied', or should the threshold be a job satisfaction level of 8—and, of course, different values will produce different estimates of the incidence of voluntary over-education. Ultimately, the need to define such thresholds is an important limitation of the analysis and it leads to some uncertainty regarding voluntary over-education estimates.

A further concern regarding the identification of voluntary over-education is the possibility individuals' job satisfaction levels and intentions to quit may change with time in the same job. Hence, individuals may move between being voluntarily and involuntarily over-educated without changing job. While such changes are highly plausible, since job attributes may change over time, a concern is that 'contentment' with jobs (i.e., individuals initially dissatisfied becoming contented due to a lowering of their expectations over time) may lead to some individuals being wrongly identified as voluntarily over-educated. Using intentions to quit to identify voluntary over-education is another potential concern. This is because such intentions are affected by individuals' labour market prospects and so, arguably, they may be inconsistent with determining whether individuals' current jobs are their preferred outcomes (i.e., individuals may report a 0 per cent chance of quitting as they

[6] Based on this distinction, individuals who accept over-educated jobs because they are associated with an increased chance of future career advancement, as predicted by career mobility theory (Sicherman and Galor, 1990; Sicherman, 1991), may also be deemed voluntarily over-educated. They are, however, different from other voluntarily over-educated individuals in two respects: first, they trade wages for the chance of future pecuniary benefits (i.e., higher wages in the future) rather than non-pecuniary benefits or improved working conditions; second, they more likely view over-education as a short-term experience. Analyses in this chapter do not attempt to distinguish between these two forms of voluntary over-education.

[7] This is not the first study to divide the over-educated into sub-groups. Chevalier (2003), then Chevalier and Lindley (2006) and Verhaest and Omey (2006b), similarly used job satisfaction levels to make a distinction, whereby those dissatisfied with the match between their education and work were deemed *genuinely over-educated* and those satisfied *apparently over-educated*. Their intention was to identify those who genuinely had under-utilised skills. In a similar vein, Green and Zhu (2010) and Mavromaras et al. (2011) used survey questions on skills utilisation to combine the over-education and over-skilling issues: Green and Zhu (2010) identified *real over-qualification* (those with under-utilised skills) and *formal over-qualification* (those who reported full utilisation of their skills), meanwhile Mavromaras et al. (2011) identified the *over-educated only*, *over-skilled only* and *over-educated and over-skilled* categories.

are unlikely to find another (better) job, but this does not mean the current job is their preferred outcome). Moreover, as discussed below, most individuals in the data report a 0 per cent chance of quitting, and so job satisfaction levels are likely driving the voluntary over-education estimates. Despite this concern, the analysis uses intentions to quit for two reasons: first, the simultaneous use of job satisfaction levels should ensure the individuals who report little chance of quitting due to poor prospects are not identified as voluntarily over-educated; second, it has the potential to improve the accuracy of the voluntary over-education estimates by helping to avoid the misclassification of the 'contented' (as such individuals are likely searching for a new job).[8] In response to these limitations and concerns regarding the identification of voluntary over-education, sensitivity analyses using various thresholds are conducted to examine the robustness of the results.

To determine appropriate thresholds for 'highly satisfied' and 'highly unlikely to quit', the distributions of overall job satisfaction levels and intentions to quit are examined for all employed individuals. These distributions are presented in Figures A7.1.1 and A7.1.2 in Appendix 7.1. The data (pooled across all eight years) reveal that approximately 70 per cent of employed individuals report a job satisfaction level in the range from 7 to 9; roughly 30 per cent report an 8 (the mode value), while the ratings of 7 and 9 each contain 20 per cent of the individuals. A further 12 per cent report a job satisfaction level of 10. Regarding intentions to quit, the data reveal that roughly 55 per cent of employed individuals report a 0 per cent chance of quitting, while at the other extreme 7 per cent report a 100 per cent chance. Further results (which are not presented, but are available from the author on request) then indicate these distributions are similar for males and females; hence, it is not necessary to define separate thresholds by gender. Ultimately, 'highly satisfied' is defined as a job satisfaction level of 8 or above and 'highly unlikely to quit' is defined as a 5 per cent chance or less of quitting.[9]

Based on these thresholds, panel A of Table 7.1 presents estimates of the incidence of voluntary over-education. In particular, the over-education estimates presented in Chapter 4 are disaggregated by whether individuals are voluntarily or involuntarily over-educated. For the eight-year period, it is estimated that roughly 6 per cent of employed males and 9 per cent of employed females are voluntarily over-educated. This means that approximately 32 per cent of over-educated males and 38 per cent of over-educated females are actually identified as voluntarily over-educated. Voluntary over-education, therefore, appears to account for a significant proportion of all the over-education in Australian labour markets.

[8] In addition, later results (Table A7.2.1 in Appendix 7.2) indicate that using intentions to quit has an effect, albeit small, on the resultant voluntary over-education estimates.

[9] This chapter does not examine the dynamics of job satisfaction levels and intentions to quit among over-educated individuals or the dynamics of voluntary and involuntary over-education; these are suitable avenues for future research.

Table 7.1: Incidence of voluntary and involuntary over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)	Overall (W1-W8)
A. Preferred estimates									
('highly satisfied' if job satisfaction ≥8; 'highly unlikely to quit' if chance quit ≤5%)									
Males									
Over-educated	*16.2*	*18.3*	*19.3*	*18.5*	*18.2*	*19.4*	*18.7*	*19.6*	*18.6*
Voluntarily	4.8	6.0	6.3	5.9	5.6	5.7	6.1	6.5	5.9
Involuntarily	11.4	12.2	13.0	12.6	12.5	13.7	12.6	13.1	12.7
N	3,141	3,006	3,034	2,952	3,060	3,127	3,112	3,165	24,597
Females									
Over-educated	*21.5*	*21.3*	*22.5*	*22.4*	*23.4*	*24.1*	*24.1*	*23.8*	*23.0*
Voluntarily	8.2	7.2	8.5	7.8	8.7	8.7	9.9	9.8	8.7
Involuntarily	13.4	14.1	14.1	14.6	14.7	15.3	14.2	14.0	14.3
N	3,077	2,864	2,904	2,865	3,060	3,131	3,167	3,190	24,258
B. Alternative estimates of voluntary over-education									
More restrictive definition (or *lower bound estimates*)									
('highly satisfied' if job satisfaction ≥9; 'highly unlikely to quit' if chance quit =0%)									
Males	2.4	3.2	3.5	2.6	3.0	2.5	3.1	2.9	2.9
Females	4.8	3.9	4.7	3.9	4.2	4.6	5.0	5.1	4.5
Less restrictive definition (or *upper bound estimates*)									
('highly satisfied' if job satisfaction ≥7; 'highly unlikely to quit' if chance quit ≤10%)									
Males	7.6	8.7	9.1	8.5	8.1	7.9	9.1	9.5	8.6
Females	10.9	10.0	11.7	11.0	11.5	12.1	12.9	13.3	11.7

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.
Italicised figures are the estimated incidences of over-education reported in Chapter 4.

Alternative thresholds are then used to examine the robustness of these estimates. Specifically, job satisfaction levels of 7, 8 and 9 are used as thresholds for 'highly satisfied' and intentions to quit of 0, 5 and 10 per cent are used as thresholds for 'highly unlikely to quit'. The results are presented in Table A7.2.1 in Appendix 7.2. They indicate that altering the 'highly unlikely to quit' threshold has a small effect; for the three values used, the estimates of voluntary over-education generally remain within a 1 percentage point range. The threshold for 'highly satisfied' has a somewhat larger effect: using a job satisfaction level of 7 (rather than 8) increases the estimates by roughly 2 percentage points, while a job satisfaction level of 9 reduces them by up to 3 percentage points. To provide an indication of the variability (or uncertainty) in the voluntary over-education estimates, panel B of Table 7.1 presents what are, arguably, lower bound and upper bound estimates.[10] The results suggest the incidence of voluntary over-education among employed males is approximately between 3 and 9 per cent, and for employed females between 4 and 12 per cent. These

[10] More restrictive thresholds for 'highly satisfied' and 'highly unlikely to quit' are used to derive the lower bound estimates, while less restrictive thresholds are used to derive upper bound estimates.

correspond to between 15 and 45 per cent of over-educated males and 18 and 50 per cent of over-educated females being voluntarily over-educated. Ultimately, despite the uncertainty associated with the estimates, the results clearly indicate that some individuals are indeed voluntarily over-educated.

7.4 Voluntary over-education and job attributes

To assess the validity of the above estimates of voluntary over-education, this section examines the relationship between such voluntary over-education and job attributes. If the identification method is valid, there should be evidence of trade-offs between wages and other job attributes among the individuals identified as voluntarily over-educated. Specifically, it should be found that, compared to well-matched individuals, the voluntarily over-educated experience wage penalties and improvements in other job attributes. The experiences of the voluntarily and involuntarily over-educated should also be different; most importantly, evidence should be consistent with the assumption that involuntary over-education is not the result of trade-offs made by individuals (i.e., it is an undesirable state). Hence, the involuntarily over-educated should be found to also incur wage penalties but without improvements in other job attributes.[11]

Econometric models are used to examine the relationship between voluntary over-education and individuals' job attributes. Such models treat job attributes as the dependent variables and include indicators for being voluntarily over-educated and involuntarily over-educated. The HILDA Survey data contain an extensive amount of information on the characteristics of each individual's current job. This chapter seeks to consider the key job attributes. In particular, the following measures are used to capture the attributes of individuals' jobs:

- real hourly wage;
- probability of job loss;
- hours (per week);
- travel time (hours per week);
- satisfaction with pay;
- satisfaction with job security;
- satisfaction with hours;

[11] A potential criticism of this validity test is that it examines measures—the job attributes—that are directly related to the key measure used to identify voluntary over-education—job satisfaction level—and so it may, by definition, lead to results that validate the voluntary over-education estimates (i.e., it is a biased empirical test). This, however, is not the case because, while the voluntarily over-educated should have better job attributes than the involuntarily over-educated (since they report higher job satisfaction levels), the differences being estimated here—those relevant to whether voluntary over-education is the result of trading wages for other job attributes—are between voluntarily over-educated and well-matched individuals (and then the involuntarily over-educated and well-matched). And, since fixed effects estimates are derived and these are essentially the result of comparing voluntarily over-educated and well-matched individuals who are otherwise identical, the voluntarily over-educated are not automatically likely to exhibit improved job attributes. Similarly, the involuntarily over-educated are not automatically likely to have different job attributes (except, of course, for an over-education wage penalty).

- satisfaction with work-life balance;
- satisfaction with the work;
- flexibility;
- autonomy;
- input;
- stressfulness;
- complexity;
- learning of new skills;
- use of existing skills;
- permanent contract;
- preferred number of hours; and,
- flexible start/finish times.[12, 13, 14]

Since job attributes are the dependent variables and they vary between being measured as continuous (e.g., real hourly wage), ordinal (e.g., satisfaction with pay) and binary (e.g., flexible start/finish times), three different econometric models and estimators must be used. Each uses a regression framework and the panel data to control for (observable and unobservable) individual heterogeneity. This ensures the resultant estimates are not confounded by other factors and reflect only the differences in job attributes that arise from being voluntarily over-educated (and involuntarily over-educated) rather than well-matched. Thus, statistically significant evidence identifies the particular job attributes for which the voluntarily over-educated trade wages.

7.4.1 Econometric models and estimators

The econometric models and estimators used to examine the continuous-, ordinal- and binary-valued job attributes are outlined below.

Continuous-valued job attributes

For continuous-valued job attributes, the underlying model for each attribute is defined as:

$$Y_{it} = X_{it}^{'}\beta + \delta_{VO}VO_{it} + \delta_{IO}IO_{it} + \alpha_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 1,...,T \qquad [7.1]$$

[12] For details on these measures (including their scales) see Appendix 7.3: Table A7.3.1 contains descriptions and details on variable derivations; Table A7.3.2 contains descriptive statistics (means and standard deviations) by over-education status.
[13] Measures using individuals' satisfaction levels are actually proxy measures for job attributes. Consider, for example, the satisfaction with job security measure: this measures an outcome associated with the job attribute (individuals' satisfaction with their level of job security) rather than the attribute itself (the level of job security). Nevertheless, these proxy measures should provide valid information on the underlying job attributes, and therefore provide relevant evidence for this chapter.
[14] Some of the typical means of non-wage compensation, such as pensions, health benefits and paid holidays, are not explored here because their receipt is essentially universal in Australia (as there is a universal health care system and legislation dictates workers' entitlements to superannuation contributions, sick leave, maternity leave and annual leave). Moreover, the minor variations that may exist across individuals cannot be adequately identified in the HILDA Survey data.

where Y_{it} are job attributes for individual i at time t, X_{it} is a vector of controls for individual characteristics at t, β is a vector of coefficients associated with X_{it}, VO_{it} and IO_{it} are indicators for being voluntarily over-educated and involuntarily over-educated at t, δ_{VO} and δ_{IO} are the estimated parameters of interest, α_i is a time-invariant random variable representing unobserved effects for individual i, and ε_{it} are idiosyncratic errors that vary across i and t (i.e., the usual regression disturbances).[15] Since α_i may contain factors correlated with both X_{it} and Y_{it} (e.g., certain personality traits being correlated with both an individual's education level and their number of hours worked per week), OLS estimation of the model in [7.1] may produce biased and inconsistent estimates of δ_{VO} and δ_{IO}. To control for such unobserved individual heterogeneity (i.e., eliminate α_i from [7.1]), the models for continuous-valued job attributes are estimated using the linear fixed effects estimator.[16, 17]

Ordinal-valued job attributes

For ordinal-valued job attributes, the underlying model for each attribute is defined as:

$$Y_{it}^{*} = X_{it}^{'}\beta + \delta_{VO}VO_{it} + \delta_{IO}IO_{it} + \alpha_i + \varepsilon_{it}, \qquad i = 1,...,N; t = 1,...,T \qquad [7.2]$$

where Y_{it}^{*} are unobserved measures of the extent individual i's job at t has certain attributes, and all other controls and parameters are as defined for [7.1].[18] The relationship between Y_{it}^{*} and the observed (ordered) Y_{it} is:

$$Y_{it} = k \text{ if } \tau_{ik} < Y_{it}^{*} \leq \tau_{ik+1}, \qquad k = 1,...,K \qquad [7.3]$$

where thresholds (or cut-points) τ_{ik} are individual-specific and strictly increasing (i.e., $\tau_{ik} < \tau_{ik+1}\ \forall k$).[19] α_i may be correlated with X_{it} and it is treated as a fixed effect, which must then be controlled for (or

---

[15] The continuous-valued job attributes examined are: real hourly wage; probability of job loss; hours; and, travel time. Treating probability of job loss as continuous, however, may not be entirely appropriate given it is bounded on the 0–100 scale. Also, analysis of the data reveals that, rather than being spread across the scale, individuals' responses tend to be concentrated at certain values (i.e., there is evidence of rounding or digit preference). In response, sensitivity analyses are performed where it is treated as ordinal-valued.

[16] In particular, models similar to [5.6] in Chapter 5 are estimated to produce fixed effects estimates. Since the fixed effects estimates for each explanatory variable are identified by the individuals whose values change over time, Table A7.4.1 in Appendix 7.4 presents information on the amount of variation in VO_{it} and IO_{it}. The results indicate that each measure contains limited *within* variation. A relevant concern then is whether the individuals whose voluntary over-education status changes (i.e., who move between the voluntarily over-educated and well-matched states) are representative of all voluntarily over-educated individuals (and similarly for the variation in involuntary over-education status). This is examined in terms of the (mean) observable characteristics of the groups (i.e., using t-tests similar to those discussed in Section 5.4.1). The results (not presented, but available from the author) indicate small differences exist: compared to all voluntarily over-educated individuals, the sample who change voluntary over-education status contains higher proportions of individuals who are younger (aged 15 to 29 years), single and without children, lower proportions of NESB migrants, and individuals with less labour market experience. Similar results are found for variation in involuntary over-education status. Such disparities may mean these fixed effects estimates are not representative of the differences in job attributes among all the voluntarily over-educated and involuntarily over-educated individuals. For further details on the fixed effects estimator see Section 5.3.2.

[17] Since the errors in [7.1] are almost certainly correlated over time for each individual, robust (panel-corrected) standard errors must be calculated to ensure valid statistical inference (Cameron and Trivedi, 2005).

[18] The ordinal-valued job attributes examined are: satisfaction with pay; satisfaction with job security; satisfaction with hours; satisfaction with work-life balance; satisfaction with the work; flexibility; autonomy; input; stressfulness; complexity; learning of new skills; and, use of existing skills.

eliminated from the model) in order to derive unbiased and consistent parameter estimates. And, conditional on X_{it} and α_i, the ε_{it} are assumed to be independently and identically (standard) logistically distributed.[20] The resultant model is a fixed effects ordered logit model, whereby the probability outcome k is observed for individual i at time t is:

$$\Pr(Y_{it} = k \mid X_{it}, \alpha_i) = \Lambda(\tau_{ik+1} - X'_{it}\beta - \alpha_i) - \Lambda(\tau_{ik} - X'_{it}\beta - \alpha_i) \qquad [7.4]$$

where $\Lambda(\cdot)$ denotes the logistic distribution function.[21]

Maximum likelihood estimation of [7.4] encounters two problems (Baetschmann, Staub and Winkelman, 2011; Jones and Schurer, 2011). First, the individual-specific thresholds (τ_{ik}) cannot be distinguished from the unobserved individual effects (α_i); instead, only $\tau_{ik} - \alpha_i \equiv \alpha_{ik}$ is identified. Second, since in this chapter, as in most empirical applications, T is fixed and relatively small, even α_{ik} cannot be estimated consistently due to the incidental parameters problem.[22] The resulting bias in α_{ik} then leads to bias in the estimates of β, which can be quite substantial (Greene, 2004). Recent studies have developed several different estimators attempting to produce consistent estimates of β. The basis of each approach is the same: Y_{it} is collapsed into a binary variable and then estimates are derived using the conditional fixed effects logit estimator (as defined by Chamberlain (1980)).

In this chapter, estimates for the above fixed effects ordered logit model are derived using the so-called BUC estimator recently developed by Baetschmann et al. (2011), where BUC stands for 'Blow-Up and Cluster' and is a description of how the estimator is implemented.[23] The approach is essentially as follows. Rather than performing a single dichotomisation of Y_{it} using a particular cut-off point k, each observation in the data is replicated K–1 times (i.e., 'Blow-Up' the data) and Y_{it} is dichotomised using each of the possible cut-off points. Based on this expanded data, the conditional fixed effects logit estimator is then used to derive the BUC estimates, with standard errors clustered at the individual-level to ensure valid statistical inference (i.e., 'Cluster' the expanded data).[24, 25]

---

[19] K corresponds to the number of categories in the observed (ordered) variables; K=11 for the satisfaction measures and K=7 for the remaining measures of job attributes.

[20] That is, if $F(\cdot)$ is used to denote the cumulative distribution function of ε_{it}, then:

$$F(\varepsilon_{it} \mid X_{it}, \alpha_i) = F(\varepsilon_{it}) = \frac{1}{1 + \exp(-\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}).$$

[21] In [7.4], the VO_{it} and IO_{it} identifiers have been included in the X_{it} vector.

[22] For discussion of the incidental parameters problem see, for example, Lancaster (2000) and Wooldridge (2003).

[23] It is also equivalent to the approach of Mukherjee, Ahn, Liu, Rathouz and Sanchez (2008) in the statistics literature.

[24] For details on the derivation of the BUC estimator and the consistency of its estimates see Baetschmann et al. (2011).

[25] There are two main alternatives for estimation of the fixed effects ordered logit model. The first estimator was developed by Das and van Soest (1999) (hereafter the DS estimator). It also uses the conditional fixed effects logit estimator with Y_{it} dichotomised at every possible cut-off point, but is a two-step procedure whereby the second step combines the estimates from all possible dichotomisations via minimum distance estimation (unlike the BUC estimator which estimates them all jointly) (Baestchmann et al., 2011). The second estimator was developed by Ferrer-i-Carbonell and Frijters (2004) (hereafter the FF estimator). It performs a single dichotomisation using an 'optimal' cut-off point determined separately for each individual (e.g., using each individual's mean value to ensure all those with time variation in Y_{it} are examined) and then uses

Coefficient estimates are reported in the results tables; these are interpreted as approximations of the marginal effect that each variable has on the latent job attribute measures—that is, the differences in job attributes that arise from being voluntarily over-educated rather than well-matched.[26, 27]

Binary-valued job attributes

For binary-valued job attributes, the underlying model for each attribute is defined as:

$$Y_{it}^{*} = X_{it}^{'}\beta + \delta_{VO}VO_{it} + \delta_{IO}IO_{it} + \alpha_i + \varepsilon_{it}, \quad i = 1,...,N; t = 1,...,T \qquad [7.5]$$

where Y_{it}^{*} are the unobservable propensities to be employed in a job with the particular attribute for individual i at t, and all other controls and parameters are as defined for [7.1].[28] Since Y_{it}^{*} are unobserved in data, the binary variables examined are:

$$Y_{it} = \begin{cases} 1 \text{ if } Y_{it}^{*} > 0 \\ 0 \text{ otherwise} \end{cases} \qquad [7.6]$$

whereby an individual is observed to have a particular job attribute when their propensity to have this attribute exceeds a certain threshold (here assumed to be zero). Once again, α_i may be correlated with X_{it} and it is treated as a fixed effect that must be controlled for in order to derive unbiased and consistent estimates of δ_{VO} and δ_{IO}. A conditional fixed effects logit or fixed effects probit estimator, therefore, appears appropriate for examining the binary-valued job attributes. The fixed effects probit estimator is preferred here because in using the conditional fixed effects logit estimator there is a significant loss of information that adversely affects the efficiency of estimates (as individuals whose job attributes do not change over time (i.e., with $Y_i = 0$ or $Y_i = 1$ in all periods) are omitted) and it is difficult to calculate partial effects estimates (as this requires information on the distribution of the unobserved effects (α_i)) (Wooldridge, 2003; Cameron and Trivedi, 2005; Jones and Schurer, 2011).[29] Also, since it is difficult to derive consistent estimates for the fixed effects probit model (due to the

---

the conditional fixed effects logit estimator. The FF estimator has been widely used in empirical studies (see, for example, Frijters, Haisken-DeNew and Shields (2004a, 2004b), D'Addio, Eriksson and Frijters (2007), Booth and van Ours (2008), Kassenboehmer and Haisken-DeNew (2009), Jones and Schurer (2011) and Schmitz (2011)). But, evidence from Monte Carlo simulations in Baestchmann et al. (2011) indicates the FF estimator is inconsistent: it produces biased estimates in finite samples, and this does not disappear with analysis of larger samples. Baestchmann et al. (2011) attributed this bias to the endogenous dichotomisation of Y_{it} (i.e., using each individual's mean value to determine their cut-off point). The BUC and DS estimators, however, are both found to be consistent estimators, though the DS estimator may produce biased estimates when examining small samples. For these reasons (and because the DS estimator is less straightforward to implement) the BUC estimator is used in this chapter.

[26] This reporting and interpretation of the estimated coefficients from fixed effects ordered logit models is common in empirical studies (see, for example, D'Addio et al. (2007), Booth and van Ours (2008), Cameron and Trivedi (2009), Baestchmann et al. (2011), Jones and Schurer (2011) and Schmitz (2011)).

[27] A further alternative for examining the ordinal-valued job attributes is to use a random effects ordered probit estimator with Mundlak-Chamberlain controls (as discussed in Chapter 6) to allow for correlation between α_i and X_{it}. Lorgelly and Lindley (2008) used such an approach (and regarded it as an approximation of a fixed effects ordered probit estimator). This estimator is subsequently used in the sensitivity analysis of results for ordinal-valued job attributes.

[28] The binary-valued job attributes examined are the indicators: permanent contract; preferred number of hours; and, flexible start/finish times.

[29] The conditional fixed effects estimator is, however, used in the sensitivity analysis of the subsequent results.

incidental parameters problem) (Wooldridge, 2003; Greene, 2004; Cameron and Trivedi, 2005), they are instead approximated using a random effects probit estimator with the Mundlak-Chamberlain approach to allowing correlation between α_i and X_{it}.[30]

Recall from Section 6.4.1, the Mundlak-Chamberlain approach allows for such correlation by parametrically specifying the relationship between α_i and X_{it} and then incorporating it into the model. This relationship is assumed to be as follows:

$$\alpha_i = \bar{X}_i' a + u_i \qquad [7.7]$$

where $\bar{X}_i$ is a vector of means for the time-varying covariates for individual *i* over the entire period examined. Incorporating [7.7] into the model in [7.5] produces:

$$Y_{it}^* = X_{it}'\beta + \delta_{VO} VO_{it} + \delta_{IO} IO_{it} + \bar{X}_i' a + u_i + \varepsilon_{it}, \quad i = 1,...,N; t = 1,...,T \qquad [7.8]$$

and since u_i is uncorrelated with X_{it}, it can be absorbed into the composite error term $v_{it} = u_i + \varepsilon_{it}$ and then a standard random effects probit estimator can be used to estimate [7.8] and derive unbiased and consistent estimates of δ_{VO} and δ_{IO}.[31] Average partial effects, rather than the estimated coefficients, are reported in the results tables as they are more readily interpretable. Specifically, they represent the percentage point differences in the likelihood of having each job attribute that result from being voluntarily over-educated rather than well-matched.[32]

Specifications and sensitivity analyses

For each of the models in [7.1], [7.2] and [7.5], X_{it} contains the same extensive set of controls for individual characteristics. These capture gender, human capital levels, recent labour market experiences, demographic characteristics and year dummies.[33] Of course, controls for time-invariant factors (i.e., gender and ethnicity) drop out of models estimated using the fixed effects estimators. For estimators using Mundlak-Chamberlain controls, the $\bar{X}_i$ are defined using all observations in the data for each individual, not just the observations for which individuals are in the restricted sample (i.e., employed, aged 15 to 64 years, not a full-time student and not self-employed). In addition, two specifications are used for each model: one estimates the differences in job attributes, compared to the well-matched, for all over-educated individuals combined (i.e., uses O_{it} rather than VO_{it} and IO_{it}); the other estimates the differences separately for the voluntarily over-educated and involuntarily

[30] An alternative is to merely assume α_i contains no such factors and use a standard logit or probit estimator; this is the approach used by McGuinness and Sloane (2011). If the no correlation assumption is violated, however, such an approach will produce biased and inconsistent estimates of the desired parameters.
[31] For further details on the random effects probit estimator with Mundlak-Chamberlain controls see Section 6.4.1.
[32] Recall from Chapter 6, the non-linearity of probit models means estimated coefficients cannot be interpreted in such a way. For further details on the derivation of average partial effects see Section 6.4.2.
[33] The complete set of controls used is listed in the Notes to Table 7.2. For further details on these explanatory variables see Appendix 5.1.

over-educated individuals (i.e., uses VO_{it} and IO_{it}). The two specifications are estimated to illustrate how evidence of voluntary over-education can be concealed in analyses, such as those performed in McGuinness and Sloane (2011) and Mavromaras et al. (2011), which merely use the first specification—that is, comparing results across the two specifications should demonstrate the importance of distinguishing between the voluntarily and involuntarily over-educated prior to examining job attributes for evidence of trade-offs. For the second specification, the distinction between voluntarily and involuntarily over-educated is based on the preferred definition of voluntary over-education—job satisfaction level of 8 or above and a 5 per cent chance or less of quitting—as presented in panel A of Table 7.1.

A series of sensitivity analyses are also performed to examine the robustness of the results. The first set uses alternative estimators. Specifically, probability of job loss is treated as ordinal rather than continuous and estimates derived using the BUC estimator.[34] Analysis of the ordinal-valued job attributes is replicated using a random effects ordered probit estimator with Mundlak-Chamberlain controls (as an approximation of a fixed effects ordered probit estimator) and a linear fixed effects estimator (where the job attributes are treated as continuous rather than ordinal). Also, analysis of the binary-valued job attributes is replicated using a conditional fixed effects logit estimator. The second set of sensitivity analyses use alternative definitions of voluntary over-education—alternative thresholds for 'highly satisfied' and 'highly unlikely to quit'—to obtain estimates; specifically, the more restrictive and less restrictive definitions from panel B of Table 7.1 are used. The empirical identification of voluntary over-education is clearly the main limitation of the analyses in this chapter. It is, therefore, important that evidence of voluntary over-education be robust to small changes in the definition (or thresholds) used. The final sensitivity analysis examines males and females separately. This relaxes the assumption that labour market behaviour does not differ by gender, and thereby considers whether there is evidence of voluntary over-education among both males and females—a relevant concern since, recall from Table 7.1, its estimated incidence is lower for males—and whether males and females are voluntarily over-educated for different reasons (i.e., trading wages for different job attributes).

7.4.2 Empirical results

As in Chapter 5, this chapter restricts attention to the sample of over-educated and well-matched individuals (i.e., observations of under-educated individuals are omitted) and analyses are performed using unbalanced panels. The data contain 43,659 observations, where, based on the estimated incidence of voluntary over-education in panel A of Table 7.1, there are 3,476 observations of

[34] In doing so, probability of job loss is converted from being measured on the 0–100 scale to a 0–10 scale (0 is a 0 per cent chance, 1 is 1–10 per cent chance, 2 is 11–20 per cent chance, and so on until 10 represents 91–100 per cent chance).

individuals voluntarily over-educated, 6,188 involuntarily over-educated and 33,995 well-matched. Table 7.2 presents the main results; panel A contains results for the continuous-valued job attributes, panel B the ordinal and panel C the binary. Also, column (I) has results for the specification that estimates differences for all over-educated individuals combined, while results in column (II) are based on the specification that distinguishes between the voluntarily and involuntarily over-educated. More specifically, column (I) reports the estimated differences in job attributes that exist between the well-matched and all over-educated individuals, while column (II) reports the differences between both the well-matched and the voluntarily over-educated and between the well-matched and the involuntarily over-educated.[35]

Recall, validating the instances of voluntary over-education requires evidence that, compared to the well-matched, the individuals identified as voluntarily over-educated experience wage penalties and improvements in other job attributes. The results in Table 7.2 indicate the voluntarily over-educated experience a wage penalty of, on average, 2.7 per cent. But, despite this wage penalty, they are more satisfied with their pay; roughly 0.6 points (on the 0–10 scale) more satisfied. It appears, therefore, the voluntarily over-educated perceive their lower wage as fair, which implies they must be receiving something in return (i.e., a preferred job attribute). This is clearly indicative of voluntary trade-offs being made between wages and other job attributes, and other results in Table 7.2 identify exactly which job attributes are being improved (or obtained). The first is job security. Voluntarily over-educated individuals report a probability of job loss that is roughly 3 percentage points lower, and they are around 0.4 points more satisfied with their job security. And this is despite being 5 percentage points less likely to be employed on a permanent contract.

A second benefit is working their preferred number of hours; the voluntarily over-educated are around 10 percentage points more likely to be working their preferred number of hours each week and their satisfaction with hours is roughly 0.8 points higher. In addition, these preferred hours are less than those being worked by well-matched individuals: approximately two hours less per week. Greater job flexibility is also obtained. In particular, voluntarily over-educated individuals are around 5 percentage points more likely to have jobs with flexible start and finish times. They also have greater flexibility to decide when to do their work and, arguably a further improvement in job flexibility, their jobs are situated closer to home, which results in, on average, around 12 minutes less travel time per week (though both of these estimated differences are not statistically significant). Given the improvements in work hours and job flexibility, the voluntarily over-educated are then better able to balance work and non-work commitments; compared to the well-matched, they are roughly 1 point more satisfied with their achieved work-life balance.

[35] Complete results from these estimations are not presented, but are available from the author on request.

Table 7.2: Estimated differences in job attributes between over-educated and well-matched individuals

	(I) Over-educated		(II) Voluntarily over-educated		(II) Involuntarily over-educated		*N*
Dependent variables	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	
A. Continuous-valued job attributes (Fixed effects estimates)							
ln(real hourly wage)	-0.052**	(0.010)	-0.027*	(0.012)	-0.064**	(0.011)	*41,698*
Probability of job loss [a]	0.182	(0.651)	-3.083**	(0.691)	1.632*	(0.698)	*41,921*
Hours (per week)	-3.018**	(0.312)	-2.293**	(0.350)	-3.340**	(0.323)	*41,984*
Travel time (hours per week)	-0.322**	(0.104)	-0.195	(0.127)	-0.380**	(0.108)	*38,072*
B. Ordinal-valued job attributes (Fixed effects ordered logit estimates)							
Satisfaction with pay [b]	-0.228**	(0.067)	0.596**	(0.081)	-0.540**	(0.071)	*129,613*
Satisfaction with job security [b]	-0.190**	(0.068)	0.401**	(0.084)	-0.407**	(0.071)	*121,407*
Satisfaction with hours [b]	-0.072	(0.065)	0.812**	(0.078)	-0.408**	(0.069)	*136,450*
Satisfaction with work-life balance [b]	0.114	(0.068)	0.967**	(0.087)	-0.185**	(0.070)	*144,797*
Satisfaction with the work [b]	-0.522**	(0.067)	0.709**	(0.080)	-0.957**	(0.070)	*115,323*
Flexibility [c]	-0.160*	(0.071)	0.076	(0.085)	-0.268**	(0.074)	*96,736*
Autonomy [c]	-0.360**	(0.070)	0.018	(0.084)	-0.522**	(0.073)	*90,569*
Input [c]	-0.415**	(0.068)	0.016	(0.081)	-0.611**	(0.071)	*92,152*
Stressfulness [c]	-0.440**	(0.068)	-0.830**	(0.082)	-0.276**	(0.072)	*93,940*
Complexity [c]	-0.788**	(0.071)	-0.735**	(0.087)	-0.811**	(0.074)	*86,986*
Learning of new skills [c]	-0.659**	(0.074)	-0.325**	(0.088)	-0.804**	(0.076)	*87,570*
Use of existing skills [c]	-0.936**	(0.074)	-0.513**	(0.089)	-1.105**	(0.077)	*74,646*
C. Binary-valued job attributes (Fixed effects probit estimates) [d]							
Permanent contract	-0.086**	(0.007)	-0.050**	(0.009)	-0.102**	(0.007)	*42,010*
Preferred number of hours	0.003	(0.010)	0.101**	(0.013)	-0.046**	(0.010)	*41,979*
Flexible start/finish times	-0.022	(0.013)	0.046**	(0.018)	-0.057**	(0.015)	*34,616*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Panels A and B report coefficient estimates; panel C reports average partial effect estimates, which are averages of the individual marginal effects that being (voluntarily or involuntarily) over-educated has on the probability of having each job attribute.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following controls: *gender*; *human capital measures* (highest education, multiple qualifications, experience, occupation tenure, employer tenure, years unemployed, English proficiency, long-term health condition); *recent labour market experiences* (number of jobs last year, proportion time employed last year, proportion time unemployed last year); *demographic characteristics* (age, ethnicity, marital status, number of children, number of resident and non-resident children by age, rest of household income, State/Territory of residence, remoteness of area of residence); and, *year dummies*. Controls for gender and ethnicity drop out of models in panels A and B, and models in panel C also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except gender and ethnicity exhibit variation over time).

(a) Probability of job loss measured from 0% (no chance) to 100% (absolute certainty).

(b) Job satisfaction levels measured on scale from 0 (totally dissatisfied) to 10 (totally satisfied).

(c) Job attributes measured on scale from 1 (strongly disagree) to 7 (strongly agree).

(d) Fixed effects probit estimates approximated using random effects probit estimator with Mundlak-Chamberlain controls.

A final area of improvement concerns the work being performed. The voluntarily over-educated have jobs roughly 0.8 points (on the 1–7 scale) less stressful than well-matched individuals. And they are around 0.7 points more satisfied with the work they perform. Another important result is that, compared to well-matched individuals, the voluntarily over-educated report their jobs use less

of their existing skills. This finding (along with the similar result in column (I) of Table 7.2) provides further evidence that the voluntarily over-educated (and over-educated individuals more generally) have human capital that is being under-utilised in their current job—and is therefore additional support for the Chapter 5 finding that over-education is not merely a statistical artefact. Voluntarily over-educated individuals also have jobs that are less complex to perform and require less learning of new skills. Such results, along with the results regarding stressfulness and skill use, suggest the voluntarily over-educated have less demanding (or 'easier') jobs. But this is not necessarily an improvement as some individuals, perhaps many, may seek jobs that are challenging to perform, jobs that make use of their existing skills and jobs that require the learning of new skills (or provide the opportunity for further training).[36]

As previously discussed, the experiences of the voluntarily and involuntarily over-educated should also be distinct, with the involuntarily over-educated incurring wage penalties but without improvements in other job attributes. Results in Table 7.2 confirm that this is indeed the case. Compared to the well-matched, involuntarily over-educated individuals experience a wage penalty of, on average, 6.4 per cent, and, furthermore, they are significantly less satisfied with their pay. Thus, it immediately appears that no voluntary (or satisfactory) trade-offs are being made between wages and other job attributes among the involuntarily over-educated. Further results confirm this. Involuntarily over-educated individuals do not experience increased job security, but instead report a higher probability of job loss and lower satisfaction with job security. They are around 5 percentage points less likely to be working their preferred number of hours, and satisfaction with hours is 0.4 points lower. They are also roughly 6 percentage points less likely to have flexible start and finish times, and have significantly less flexibility in deciding when they work. Without their preferred hours and job flexibility, the involuntarily over-educated are then less satisfied with their achieved work-life balance. Finally, while they are similar to the voluntarily over-educated and work in less demanding jobs, there are important differences. In particular, they have less autonomy and input at work, experience a smaller reduction in stressfulness (roughly a 0.3 point reduction compared to the 0.8 point reduction experienced by the voluntarily over-educated) and they are significantly less satisfied with the work they perform. A further difference is that the involuntarily over-educated report a greater under-utilisation of their skills (roughly a 1.1 point difference compared to the 0.5 difference among the voluntarily over-educated). Given the relationship between human capital utilisation and wages, this difference then, at least partly, explains the greater wage penalty incurred by the involuntarily over-

[36] The issue of whether individuals perceive complexity, learning and skills use as desirable job attributes could, of course, be further examined. But, since the results have already provided evidence of the voluntarily over-educated having improved job attributes, such analyses are not been performed here. This issue is a suitable avenue for future research.

educated.[37] Ultimately, these results clearly indicate significant differences in the experiences of voluntarily over-educated and involuntarily over-educated individuals.

Comparing results across the two specifications it is also clear that use of a basic specification—as in column (I)—can fail to identify evidence of voluntary over-education. That is, the results in column (II) indicate that there are many job attributes where (compared to well-matched individuals) the voluntarily and involuntarily over-educated have statistically significant differences that operate in opposite directions. Specifically, the differences in satisfaction with pay, probability of job loss, satisfaction with job security, likelihood of working preferred number of hours, satisfaction with hours, likelihood of having flexible start and finish times, satisfaction with work-life balance and satisfaction with the work being performed. But, since over-education appears involuntary for the majority of individuals, the estimated differences in column (I) tend to reflect the circumstances of the involuntarily over-educated. In examining job attributes for evidence of voluntary over-education, therefore, it is critical to first distinguish between voluntarily and involuntarily over-educated individuals.

Results from the sensitivity analyses using alternative estimators, which are presented in Table A7.4.2 in Appendix 7.4, confirm the previous findings.[38] In particular, while the magnitudes of the estimates differ, the results are qualitatively identical to those in Table 7.2. The above evidence of voluntary over-education therefore appears robust to the choice of estimators used to examine individuals' job attributes.[39] Table 7.3 presents results from the sensitivity analyses using alternative definitions of voluntary over-education; column (I) the more restrictive definition and column (II) the less restrictive definition. Based on the more restrictive definition, the wage penalty incurred by voluntarily over-educated individuals is 1.7 per cent, which is smaller than that found in Table 7.2 and not statistically significant.[40] For this group of voluntarily over-educated individuals, the

[37] The greater wage penalty of the involuntarily over-educated may also reflect differences in innate ability levels. Specifically, it may be evidence that individuals involuntarily over-educated have lower innate ability levels than those voluntarily over-educated. This may then also be an important factor in the incidence of involuntary over-education and is, therefore, a suitable avenue for future research.

[38] To be clear, panel A of Table A7.4.2 contains the results from treating probability of job loss as ordinal and using the BUC estimator, panel B contains the random effects ordered probit estimates for the ordinal-valued job attributes, panel C contains the results from treating the ordinal-valued job attributes as continuous and using the linear fixed effects estimator and panel D contains the conditional fixed effects logit estimates for the binary-valued job attributes.

[39] Similar evidence was also derived in a separate analysis of changes in job attributes for individuals observed entering voluntary over-education. Specifically, based on regression analyses and compared to those who remained well-matched in adjacent waves of the data, individuals observed moving from well-matched to voluntarily over-educated were found to incur wage reductions, but then also experienced improvements in other job attributes. For example, and similar to above, they experienced increased job security, an increased likelihood of working their preferred number of hours, an increased likelihood of having flexible start and finish times, increased satisfaction with work-life balance and increased satisfaction with the work being performed. And, once again, involuntary over-education was a different state: individuals who moved from well-matched to involuntarily over-educated incurred (larger) wage reductions, but without improvements in other job attributes. The results from this supplementary analysis are not presented, but are available from the author on request.

[40] Despite the statistically insignificant wage penalty, these individuals still appear to be over-educated because, compared to otherwise identical well-matched individuals, they report their jobs use less of their existing skills.

estimated difference in their satisfaction with pay is then larger: they are roughly 1 point more satisfied, compared to the previous 0.6 points. In fact, with the exception of job security, these voluntarily over-educated individuals experience greater improvements in all job attributes (i.e., the absolute values of the estimated differences are larger than the corresponding figures in Table 7.2).[41] Such findings are entirely sensible: these voluntarily over-educated individuals should have greater improvements to job attributes because, compared to the voluntarily over-educated in Table 7.2, they are more satisfied with their trade-offs between wages and other job attributes (as shown by their job satisfaction levels of 9 or 10, 0 per cent chance of quitting and higher satisfaction with pay). For the involuntarily over-educated identified using the more restrictive definition, the magnitudes of the estimated differences in job attributes are all reduced compared to those in Table 7.2. This is also as expected, because the group includes individuals previously categorised as voluntarily over-educated who were found (in Table 7.2) to experience improvements in job attributes; hence, their improvements are offsetting the typically worsened job attributes of involuntarily over-educated individuals. Nevertheless, the evidence continues to indicate voluntary and involuntary over-education are distinct states.

Based on the less restrictive definition, the wage penalty incurred by voluntarily over-educated individuals is, on average, 3 per cent. Then, in contrast to above, their estimated differences in job attributes are generally reduced in magnitude—compared to the voluntarily over-educated in Table 7.2, these individuals experience smaller improvements in job attributes. On the other hand, the estimated differences in job attributes for the involuntarily over-educated are generally increased in magnitude. These findings are, once again, in line with expectations: compared to Table 7.2, the group identified as voluntarily over-educated includes individuals who are less satisfied with their job attributes (i.e., those with a job satisfaction level of 7 and a 5 to 10 per cent chance of quitting) and, as a result, they are likely to have smaller improvements in job attributes. It is also possible that some experience no such improvements and are therefore wrongly identified as voluntarily over-educated—as discussed in Section 7.3, this may be the case for individuals who become 'contented' with their jobs over time. Regardless, adding these individuals should—and does—diminish the estimated differences in job attributes for voluntarily over-educated individuals found in Table 7.2. For the involuntarily over-educated, a less restrictive definition increases the magnitudes of the (negative) estimated differences in job attributes because this group now comprises individuals who are less satisfied with their job attributes and, therefore, likely to have less favourable job attributes.

[41] These individuals still experience increased job security (compared to the well-matched), but, compared to the voluntarily over-educated in Table 7.2, their reduced probability of job loss is not as large (a 2.7 percentage point reduction compared to 3.1 in Table 7.2). This difference is relatively minor, especially given these voluntarily over-educated individuals have a much larger improvement in their satisfaction with job security (0.7 points compared to 0.4 in Table 7.2).

Table 7.3: Estimated differences in job attributes between over-educated and well-matched individuals—Using alternative thresholds to identify voluntary over-education

	(I) More restrictive definition (Job satisfaction≥9; chance quit=0%)				(II) Less restrictive definition (Job satisfaction≥7; chance quit≤10%)			
	Voluntarily over-educated		Involuntarily over-educated		Voluntarily over-educated		Involuntarily over-educated	
Dependent variables	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*
A. Continuous-valued job attributes (Fixed effects estimates)								
ln(real hourly wage)	-0.017	(0.014)	-0.059**	(0.010)	-0.037**	(0.011)	-0.064**	(0.011)
Probability of job loss [a]	-2.714**	(0.787)	0.697	(0.666)	-2.763**	(0.678)	2.509**	(0.725)
Hours (per week)	-2.460**	(0.393)	-3.117**	(0.318)	-2.313**	(0.338)	-3.574**	(0.328)
Travel time (hours per week)	-0.272	(0.145)	-0.331**	(0.105)	-0.261*	(0.119)	-0.371**	(0.113)
B. Ordinal-valued job attributes (Fixed effects ordered logit est.s)								
Satisfaction with pay [b]	0.978**	(0.097)	-0.415**	(0.069)	0.362**	(0.076)	-0.620**	(0.072)
Satisfaction with job security [b]	0.739**	(0.101)	-0.330**	(0.069)	0.239**	(0.077)	-0.467**	(0.073)
Satisfaction with hours [b]	1.259**	(0.096)	-0.266**	(0.067)	0.530**	(0.073)	-0.473**	(0.071)
Satisfaction work-life balance [b]	1.342**	(0.107)	-0.053	(0.070)	0.680**	(0.079)	-0.246**	(0.073)
Satisfaction with the work [b]	1.107**	(0.103)	-0.754**	(0.068)	0.397**	(0.074)	-1.093**	(0.073)
Flexibility [c]	0.154	(0.099)	-0.219**	(0.072)	-0.014	(0.079)	-0.279**	(0.077)
Autonomy [c]	0.070	(0.096)	-0.436**	(0.071)	-0.114	(0.077)	-0.542**	(0.075)
Input [c]	0.064	(0.096)	-0.508**	(0.069)	-0.086	(0.076)	-0.669**	(0.073)
Stressfulness [c]	-1.015**	(0.099)	-0.339**	(0.070)	-0.765**	(0.077)	-0.197**	(0.074)
Complexity [c]	-0.668**	(0.102)	-0.812**	(0.072)	-0.692**	(0.080)	-0.864**	(0.076)
Learning of new skills [c]	-0.344**	(0.101)	-0.719**	(0.074)	-0.354**	(0.082)	-0.891**	(0.079)
Use of existing skills [c]	-0.505**	(0.104)	-1.009**	(0.075)	-0.575**	(0.082)	-1.192**	(0.079)
C. Binary-valued job attributes (Fixed effects probit estimates) [d]								
Permanent contract	-0.042**	(0.012)	-0.094**	(0.007)	-0.051**	(0.008)	-0.113**	(0.008)
Preferred number of hours	0.120**	(0.017)	-0.021*	(0.010)	0.078**	(0.012)	-0.064**	(0.011)
Flexible start/finish times	0.077**	(0.022)	-0.044**	(0.014)	0.027	(0.016)	-0.066**	(0.016)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Panels A and B report coefficient estimates; panel C reports average partial effect estimates, which are averages of the individual marginal effects that being (voluntarily or involuntarily) over-educated has on the probability of having each job attribute.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *gender*; *human capital measures*; *recent labour market experiences*; *demographic characteristics*; and, *year dummies*. See Notes to Table 7.2 for exact list of variables contained in these categories. Controls for gender and ethnicity drop out of models in panels A and B, and models in panel C also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except gender and ethnicity exhibit variation over time).

(a) Probability of job loss measured from 0% (no chance) to 100% (absolute certainty).

(b) Job satisfaction levels measured on scale from 0 (totally dissatisfied) to 10 (totally satisfied).

(c) Job attributes measured on scale from 1 (strongly disagree) to 7 (strongly agree).

(d) Fixed effects probit estimates approximated using random effects probit estimator with Mundlak-Chamberlain controls.

Despite the changes in magnitudes discussed above, the results in Table 7.3 are entirely consistent with the evidence of voluntary over-education found in Table 7.2. Specifically, individuals identified as voluntarily over-educated are still found to incur wage penalties and, in exchange, experience improvements in other job attributes. And the evidence continues to support the distinction between voluntary and involuntary over-education. Evidence validating the existence of

voluntary over-education is, therefore, robust to small changes in the definition (or thresholds) used to identify voluntary over-education.[42]

Table 7.4 presents results from the final sensitivity analysis, which examines males and females separately. The evidence again indicates individuals identified as voluntarily over-educated experience trade-offs between wages and other job attributes, and the experiences of the voluntarily and involuntarily over-educated remain distinct. An exception to this, however, is that the estimated wage penalty for voluntarily over-educated males is only 1 per cent and not statistically significant. This may mean no such voluntary over-education exists among males because no wage penalty means no trade-offs. But, further evidence in Table 7.4 arguably suggests that they incur a small wage penalty, rather than no wage penalty. In particular, compared to voluntarily over-educated females, who incur a 4.7 per cent (and statistically significant) wage penalty, these males report a greater utilisation of their human capital in their current job. Voluntarily over-educated males, therefore, incur a small wage penalty because there is only a minor under-utilisation of their human capital.

The results in Table 7.4 also suggest that males and females are voluntarily over-educated for different reasons, with some of the improvements in job attributes observed in Table 7.2 being predominantly driven by the experiences of females. Specifically, the results indicate the following. Voluntarily over-educated males trade wages for: increased job security, with roughly a 3.5 percentage point reduction in the probability of job loss; the ability to work their preferred number of hours, which is around 1.7 hours less per week (than well-matched males); improved work-life balance (which, given there is no evidence of reduced travel time or an increased likelihood of having flexible start and finish times, is likely due to working fewer hours); and, greater satisfaction with the work being performed, as reflected by increased autonomy and input at work and a job that is less stressful and difficult to perform (though some of these estimated differences are not statistically significant). Voluntarily over-educated females, on the other hand, trade wages for: increased job security in the order of a 2.7 percentage point reduction in the probability of job loss; preferred number of work hours, which are roughly 2.6 hours less per week (than well-matched females); greater job flexibility, in the form of flexible start and finish times and less travel time to work; improved work-life balance (which is likely due to the reduced hours and greater job flexibility); and, less demanding jobs, which are less stressful, less complex to perform, require less learning of new skills and use less of their existing skills.

[42] Since estimates based on the less restrictive definition appear valid, it could be argued that this should be used as the basis for the preferred estimates of voluntary over-education. Recall from Table 7.2, it would then be estimated that around 9 per cent of employed males and 12 per cent of employed females are voluntarily over-educated; meaning roughly half of all over-educated males and females are actually voluntarily over-educated. At the very least, these appear to be appropriate upper bound estimates for the incidence of voluntary over-education.

Table 7.4: Estimated differences in job attributes between over-educated and well-matched individuals—Analyses by gender

	(I) Males				(II) Females			
	Voluntarily over-educated		Involuntarily over-educated		Voluntarily over-educated		Involuntarily over-educated	
Dependent variables	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*	*Coeff. / APE*	*Robust SE*
A. Continuous-valued job attributes (Fixed effects estimates)								
ln(real hourly wage)	-0.010	(0.011)	-0.058**	(0.015)	-0.047**	(0.016)	-0.069**	(0.015)
Probability of job loss [a]	-3.481**	(1.064)	1.184	(1.028)	-2.706**	(0.911)	2.055*	(0.946)
Hours (per week)	-1.732**	(0.480)	-2.637**	(0.424)	-2.560**	(0.492)	-3.799**	(0.466)
Travel time (hours per week)	-0.026	(0.189)	-0.250	(0.158)	-0.280	(0.170)	-0.465**	(0.149)
B. Ordinal-valued job attributes (Fixed effects ordered logit est.s)								
Satisfaction with pay [b]	0.634**	(0.124)	-0.684**	(0.106)	0.612**	(0.108)	-0.413**	(0.094)
Satisfaction with job security [b]	0.421**	(0.128)	-0.407**	(0.104)	0.392**	(0.112)	-0.415**	(0.096)
Satisfaction with hours [b]	0.714**	(0.121)	-0.522**	(0.102)	0.892**	(0.103)	-0.325**	(0.093)
Satisfaction work-life balance [b]	0.865**	(0.135)	-0.243*	(0.104)	1.047**	(0.115)	-0.138	(0.097)
Satisfaction with the work [b]	0.778**	(0.121)	-0.861**	(0.102)	0.660**	(0.108)	-1.036**	(0.097)
Flexibility [c]	0.134	(0.134)	-0.239*	(0.110)	0.027	(0.111)	-0.298**	(0.100)
Autonomy [c]	0.155	(0.134)	-0.418**	(0.111)	-0.057	(0.107)	-0.589**	(0.097)
Input [c]	0.164	(0.120)	-0.607**	(0.103)	-0.055	(0.109)	-0.596**	(0.097)
Stressfulness [c]	-0.597**	(0.126)	-0.186	(0.106)	-1.010**	(0.110)	-0.357**	(0.098)
Complexity [c]	-0.614**	(0.133)	-0.738**	(0.109)	-0.823**	(0.118)	-0.858**	(0.102)
Learning of new skills [c]	-0.028	(0.130)	-0.552**	(0.111)	-0.549**	(0.120)	-0.995**	(0.103)
Use of existing skills [c]	-0.323*	(0.135)	-0.874**	(0.111)	-0.683**	(0.120)	-1.321**	(0.108)
C. Binary-valued job attributes (Fixed effects probit estimates) [d]								
Permanent contract	-0.022	(0.012)	-0.083**	(0.009)	-0.057**	(0.018)	-0.114**	(0.011)
Preferred number of hours	0.118**	(0.021)	-0.044**	(0.016)	0.090**	(0.017)	-0.047**	(0.014)
Flexible start/finish times	-0.041	(0.026)	-0.110**	(0.020)	0.110**	(0.023)	-0.008	(0.020)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Panels A and B report coefficient estimates; panel C reports average partial effect estimates, which are averages of the individual marginal effects that being (voluntarily or involuntarily) over-educated has on the probability of having each job attribute. Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *human capital measures*; *recent labour market experiences*; *demographic characteristics*; and, *year dummies*. See Notes to Table 7.2 for exact list of variables contained in these categories. Controls for ethnicity drop out of models in panels A and B, and models in panel C also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).

(a) Probability of job loss measured from 0% (no chance) to 100% (absolute certainty).

(b) Job satisfaction levels measured on scale from 0 (totally dissatisfied) to 10 (totally satisfied).

(c) Job attributes measured on scale from 1 (strongly disagree) to 7 (strongly agree).

(d) Fixed effects probit estimates approximated using random effects probit estimator with Mundlak-Chamberlain controls.

Based on this comparison, males appear more likely to be voluntarily over-educated because they are seeking increased job security, fewer work hours (though less of a reduction than females) and greater autonomy and input at work, while females seek increased job security (though less so than males), fewer work hours, greater job flexibility and less demanding jobs. Moreover, the evidence in Table 7.2 of an increased likelihood of flexible start and finish times, reduced travel time and reduced need to learn new skills is driven by the experiences of females. And, to a lesser extent,

so is the evidence of reduced stressfulness and reduced skills utilisation. These gender differences may reflect the typically greater attachment to the labour force among males, and the division of household and child-rearing duties between males and females that results in females requiring greater flexibility in their jobs and less hours of work to accommodate such duties.

7.5 Persistence and voluntary over-education

Given the preceding evidence of voluntarily over-educated individuals, this chapter concludes by (briefly) considering the supplementary issue of whether such voluntary over-education explains the persistence in over-education (as observed in Chapter 6). That is, it examines the possibility that evidence of persistent over-education arises because such voluntarily over-educated individuals have achieved their preferred outcome and, as a result, choose to remain in these (over-educated) jobs over time. If indeed the case, it would mean that persistent over-education is not necessarily a sign of labour market failure and that not all persistently over-educated individuals should be targeted by government policy interventions aiming to prevent and resolve over-education. Instead, the suitable target for such policy interventions would be persistently and involuntarily over-educated individuals. To consider this issue, the correlation between persistent and voluntary over-education is examined. Table 7.5 presents the results. In particular, it contains the incidences of over-education in each year (as reported in Chapter 4), the proportions of these incidences that represent individuals persistently over-educated over the eight-year period (as reported in Chapter 6), and the proportions of these persistently over-educated who are voluntarily over-educated in that particular year (as identified in this chapter using the preferred definition of voluntary over-education).

Table 7.5: Persistent over-education by voluntariness (at time *t*) and gender (%)

	Year (wave) which corresponds to time *t*							
	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Males								
Over-educated at t	*16.2*	*18.3*	*19.3*	*18.5*	*18.2*	*19.4*	*18.7*	*19.6*
% persistently over-educated	40.4	41.4	39.7	41.8	41.3	40.8	40.6	36.1
% voluntarily over-educated at *t*	42.9	38.6	44.0	46.5	41.1	35.1	44.7	41.6
N	298	302	322	308	305	320	317	335
Females								
Over-educated at t	*21.5*	*21.3*	*22.5*	*22.4*	*23.4*	*24.1*	*24.1*	*23.8*
% persistently over-educated	36.8	39.5	35.4	32.2	32.7	30.1	31.0	29.0
% voluntarily over-educated at *t*	49.7	45.3	46.6	53.1	44.4	46.9	52.3	49.1
N	390	373	417	414	430	458	449	459

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on the sample of individuals in the balanced panel; N refers to the number over-educated at *t*. Italicised figures are the estimated incidences of over-education reported in Chapter 4. Individuals considered *persistently over-educated* if they are observed over-educated in 7 or 8 waves of the data; proportions above correspond to the figures reported in Table 6.2 in Chapter 6. Individuals identified as *voluntarily over-educated* using the preferred definition of voluntary over-education, as presented in panel A of Table 7.1.

The evidence indicates voluntary over-education explains much of the observed persistence in over-education. As an example, consider the results for males in 2001: among the 16.2 per cent over-educated in 2001, 40.4 per cent are persistently over-educated over the entire period, of which 42.9 per cent are voluntarily over-educated in 2001. For females, this figure is 49.7 per cent. Generally, around 42 per cent of persistently over-educated males and 48 per cent of persistently over-educated females are voluntarily over-educated.[43] Almost half of all persistently over-educated individuals, therefore, are actually voluntarily over-educated.

7.6 Discussion and conclusion

The aim of this chapter was to consider the possibility that some individuals are voluntarily over-educated. It was assumed over-educated individuals who are both highly satisfied with their job and highly unlikely to quit are voluntarily over-educated, while the remainder are involuntarily over-educated. To validate this distinction and the resultant estimates, the relationship between such voluntary over-education and job attributes was examined. A series of job attributes was considered and empirical estimators used to estimate the differences in job attributes between voluntarily over-educated and well-matched individuals, and the differences between the involuntarily over-educated and well-matched. It was assumed evidence of trade-offs between wages and other job attributes among the voluntarily over-educated, but not the involuntarily over-educated, would validate the empirical identification of voluntary over-education.

Given assumptions as to what represents 'highly satisfied' and 'highly unlikely to quit', it was estimated that roughly 6 per cent of employed males and 9 per cent of employed females are voluntarily over-educated. Thus, around 32 per cent of over-educated males and 38 per cent of over-educated females are actually voluntarily over-educated. It was also found that voluntarily over-educated individuals incur wage penalties of, on average, roughly 3 per cent and that they experience improvements in a host of other job attributes. In particular, they trade wages for: increased job security; working their preferred number of hours per week (which are fewer than those worked by the well-matched); greater job flexibility, in the form of flexible start and finish times, flexibility to decide when to work and a job closer to home; and, greater satisfaction with the work being performed, which appears to result from working in a less demanding job (i.e., a job which is less stressful, less complex to perform, requires less learning of new skills and uses less of their existing skills). And, seemingly as a culmination of the preceding factors, the voluntarily over-educated are significantly more satisfied with their achieved work-life balance than well-matched individuals.

[43] Sensitivity analyses using duration over-educated to define persistence (which was also done in Chapter 6 and leads to larger sample sizes for these analyses) produce similar results; these results are presented in Table A7.5.1 in Appendix 7.5.

Meanwhile, the involuntarily over-educated incur wage penalties of, on average, roughly 6 per cent, but experience no such improvements in job attributes. This chapter, therefore, has found evidence that some individuals are indeed voluntarily over-educated.

Since it arises due to individuals' preferences regarding job attributes, these instances of voluntary over-education should not be considered the result of labour market failures, at least from the perspective of individuals. This means the general perception in existing research that all over-education represents labour market failures must be modified: henceforth, only instances in which individuals are involuntarily over-educated should be perceived in such a way. The possibility of voluntary over-education also means that persistence in over-education, as observed in Chapter 6, need not always represent an undesirable outcome. In fact, supplementary analyses in this chapter found almost half of all persistently over-educated individuals fit the definition of voluntarily over-educated. As a result, it appears government policy interventions cannot simply be targeted at all persistently over-educated individuals collectively. Moreover, based on these findings, it appears that involuntarily over-educated individuals should be the target of government policy interventions to prevent and resolve over-education, particularly those individuals who have also been (or have the potential to be) persistently over-educated.

For the voluntarily over-educated, it appears no such interventions are necessary because they would result in welfare losses for individuals. From the perspective of societies, however, voluntary over-education is arguably still representative of labour market failures given it is associated with the under-utilisation of the human capital available in the workforce (i.e., an inefficient allocation of resources). Hence, welfare gains—higher productivity levels, economic growth rates and living standards—would be achieved from government policy interventions that can, without adversely affecting the individuals, lead the voluntarily over-educated to instead accept well-matched jobs. Evidence derived in this chapter suggests that such interventions should be aimed at improving the non-pecuniary benefits and working conditions of relevant jobs (i.e., making (suitable) well-matched jobs more attractive to the voluntarily over-educated). This means adjusting or creating jobs such that they utilise the education of the voluntarily over-educated and have the attributes desired by these individuals—specifically, the increased job security and flexibility regarding hours worked that enables them to achieve their preferred work-life balance. And, for the realisation of welfare gains, this re-allocation of the voluntarily over-educated must not displace already well-matched individuals. Given these policy implications, there is a need for further research into the characteristics and circumstances of voluntarily and involuntarily over-educated individuals.

Recall, the over-education literature has focused on wages as the only benefit to education investments and thereby overlooked the role individuals' preferences play in the incidence of over-education. This chapter, therefore, has made several contributions to the literature. Specifically, it has

provided empirical evidence that some over-education arises due to individuals' preferences regarding job attributes: some individuals are *voluntarily* over-educated. In doing so, it has provided the first estimates of the incidence of such voluntary over-education, along with evidence on the particular job attributes for which these voluntarily over-educated individuals trade wages—their reasons for accepting over-educated jobs. It has also provided evidence that voluntary over-education explains much of the observed persistence in over-education.

The empirical test performed in this chapter has some limitations. The main limitation is the somewhat arbitrary definition of thresholds for 'highly satisfied' and 'highly unlikely to quit', which leads to some uncertainty regarding the voluntary over-education estimates. Sensitivity analyses, however, derived lower and upper bound estimates—which suggest incidences between 3 and 9 per cent for employed males and between 4 and 12 per cent for employed females—and these continued to provide evidence that some individuals are voluntarily over-educated (as the subsequent analysis of job attributes also validated these estimates). Further regarding these thresholds, this chapter did not consider the dynamics of job satisfaction levels and intentions to quit among over-educated individuals, or how these may affect the identification of voluntary over-education (e.g., evidence of over-educated individuals becoming 'contented' with their jobs over time). Thus, examining such dynamics and the dynamics of voluntary and involuntary over-education are suitable avenues for future research. Also, since most employed individuals report a 0 per cent chance of quitting, the voluntary over-education estimates—the distinctions between voluntarily and involuntarily over-educated—are largely determined by individuals' job satisfaction levels.

The empirical test also adopts a static approach to labour markets as it assumed evidence of *contemporaneous* trade-offs—wage penalties and improvements in other job attributes experienced in the same time period—validates the existence of voluntary over-education. This is a simplification because, in reality, individuals are likely to make employment decisions with respect to maximising their *lifetime* utility levels. A further limitation, as in analyses in previous chapters, is that the estimated differences in job attributes that have been derived represent merely the *average* differences for all voluntarily (and involuntarily) over-educated individuals. Hence, the evidence does not necessarily guarantee that all the individuals identified as voluntarily over-educated experience a wage penalty and improvements in other job attributes. The empirical test is also reliant on the assumption that these estimated differences are not confounded by other factors—that is, they reflect only the differences in job attributes that arise from being voluntarily (and involuntarily) over-educated rather than well-matched. This may be prevented by the fact that the fixed effects probit estimates are approximated using a random effects probit estimator with Mundlak-Chamberlain controls and that the linear fixed effects estimates may not be representative of the differences among all the voluntarily (and involuntarily) over-educated individuals (as the sub-samples that identify them are

not entirely representative of all voluntarily (and involuntarily) over-educated individuals). Sensitivity analyses using alternative estimators, however, produce qualitatively similar results.

Chapter 8

Conclusions

This study has investigated the existence of over-education in Australian labour markets. Based on analysis of individual-level data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey for the 2001 to 2008 period, empirical evidence has been found demonstrating that such over-education does indeed exist. In particular, it was estimated that approximately 20 per cent of working-age employees are over-educated in each year. An empirical test then validated these estimates: the individuals identified as over-educated were found to have human capital from education that is under-utilised in their current job. This empirical test considered the relationship between over-education and individuals' wages, and it was found that being over-educated causes an individual's wage to be lower than if they were instead well-matched. The size of this over-education wage penalty is, on average, approximately 9 per cent for males and 7 per cent for females.

Over-education was also found to be persistent for a sizeable group of individuals. Generally, 30 to 40 per cent of the over-education in each year corresponds to individuals who are persistently over-educated, with the majority of these individuals having been over-educated for in excess of ten years. Empirical evidence confirmed that those identified as persistently over-educated still have under-utilised human capital (i.e., remain over-educated even after such extended periods). In particular, the over-education wage penalty was found to be relatively constant over time spent over-educated. Empirical evidence that over-education can lead to human capital depreciation was also found. Specifically, state dependence in over-education was found: being over-educated in the previous year increases the likelihood an individual is currently over-educated by approximately 31 percentage points for males and 35 percentage points for females (or it makes them approximately six times more likely to be currently over-educated). Moreover, for approximately 40 per cent of the individuals observed persistently over-educated, it is state dependence that causes this persistence. Hence, it is concluded that instances of over-education are not merely short-term labour market disequilibria that have no enduring effects.

Over-education was then found to be voluntary for some individuals. In particular, it was estimated that approximately 6 per cent of employed males and 9 per cent of employed females—or approximately 32 per cent of the over-educated males and 38 per cent of the over-educated females—are voluntarily over-educated. It was found that these individuals incur wage penalties of, on average, approximately 3 per cent and, in exchange, they experience improvements in other job

attributes, including: increased job security; working their preferred number of hours per week; greater job flexibility (e.g., flexible start and finish times); and, greater satisfaction with the work being performed (which appears to result from working in a less demanding or less stressful job). The voluntarily over-educated were also found to be significantly more satisfied with their achieved work-life balance (compared to otherwise identical well-matched individuals), and this is seemingly the culmination of the preceding factors. Meanwhile, it was found that the individuals deemed involuntarily over-educated incur wage penalties of, on average, approximately 6 per cent, but they experience no such improvements in job attributes.

This study, therefore, has found evidence of over-education in Australian labour markets. It has found instances of over-education are not merely a result of the inherently dynamic nature of labour markets; instead, some represent labour market failures that require interventions to be resolved. And it has found some over-educated individuals have actually achieved their preferred outcome; hence, not all instances of over-education are necessarily the result of labour market failures. Nevertheless, this study has found evidence of labour market failures—inefficiencies in the matching of individuals and jobs—that are leading to an under-utilisation of the human capital individuals derive from education in Australia. These instances of over-education represent a source of potential gains for the economy: having over-educated individuals instead well-matched would result in higher wages for the individuals and then higher productivity levels, economic growth rates and living standards. Government policy interventions to prevent and resolve such over-education could, therefore, benefit the Australian economy.

Furthermore, given the evidence that over-education can be persistent, temporary, voluntary and involuntary for individuals, such government policy interventions should be tailored specifically for different groups of individuals (rather than being targeted at all over-educated individuals collectively). Since over-education can have enduring effects (i.e., can lead to human capital depreciation), it also appears interventions to prevent over-education are of greater importance. Interventions aimed at persistent over-education are likely to result in the greatest benefit for the economy, though it must remembered that almost half of all persistently over-educated individuals fit the definition of voluntarily over-educated. For the voluntarily over-educated, it appears no interventions are necessary because they would result in welfare losses for individuals. From the perspective of society, however, voluntary over-education is arguably still representative of labour market failures given it is associated with the under-utilisation of the human capital available in the workforce (i.e., an inefficient allocation of resources). Hence, welfare gains—higher productivity levels, economic growth rates and living standards—would be achieved from interventions that can, without adversely affecting the individuals, lead the voluntarily over-educated to instead accept well-matched jobs. Evidence derived here suggests that such interventions should be aimed at improving

the non-pecuniary benefits and working conditions of relevant jobs (i.e., making these well-matched jobs more attractive to the voluntarily over-educated). The realisation of welfare gains from any government policy interventions depends on three key factors. First, well-matched jobs must either already exist or can be created for these over-educated individuals. The re-allocation of over-educated individuals must not displace those already well-matched. And, of course, the cost of such interventions affects whether welfare gains are realised. Ultimately, rather than suggesting a widespread over-investment in education in Australia, this study has highlighted that the extent to which individuals and society benefit from investments in education is dependent on the quality of the individual-job matches, with respect to human capital utilisation, which are achieved in Australian labour markets.

Prior to this study, over-education research had several important limitations. In particular, there were few studies that empirically tested the validity of over-education and its identification, and, due to endogeneity bias and data limitations in previous studies, there was no clear evidence on whether over-education has a causal effect on individuals' wages. There was little empirical evidence regarding the dynamics of over-education. And the role individuals' preferences play in the incidence of over-education had been overlooked. This study, therefore, has made several contributions to the over-education literature. Specifically, Chapter 5 has provided empirical evidence that validates the existence of over-education in labour markets. It has provided over-education wage penalty estimates that can be interpreted as causal effects (i.e., found over-education has a negative causal effect on individuals' wages). Moreover, it found that pooled OLS estimates, as reported in most previous studies, significantly over-estimate the size of this over-education wage penalty. Chapter 6 has provided empirical evidence that over-education is persistent for a sizeable group of individuals. It has provided evidence on the dynamics of the over-education wage penalty, both for individuals who remain over-educated over time and those who exit to well-matched employment. It has also provided evidence of state dependence in over-education, which is considered evidence of a link between over-education and human capital depreciation. Thus, it found that a dynamic view of labour markets does not diminish the importance of over-education as labour market failures. Chapter 7 has provided empirical evidence that some over-education arises due to individuals' preferences regarding job attributes: some individuals are voluntarily over-educated. It has provided estimates of the incidence of such voluntary over-education, along with evidence on the particular job attributes for which these individuals trade wages. It has also provided evidence that voluntary over-education explains much of the observed persistence in over-education.

Despite the contributions of this study, there remain several key avenues for future research regarding over-education. The first arises from the limitations that are associated with the empirical identification of over-education. In particular, future research should consider the potential for

individuals' innate abilities, quality of educational institution attended and qualification vintage to affect the human capital derived from education, and then whether such factors affect the likelihood of being over-educated. Moreover, the definition of over-education should be extended to consider whether individuals having multiple qualifications or whether mismatches with respect to subject matter (or field of study) also lead to an under-utilisation of the human capital individuals derive from education. Future research should also further examine the heterogeneity in the over-education wage penalty, individuals' durations over-educated and the effects of persistent over-education (e.g., whether the wage penalty, job satisfaction, life satisfaction and intentions to quit vary by duration over-educated). Further examination of the link between over-education and human capital depreciation is also warranted. Evidence individuals' human capital depreciates while over-educated then affects the identification and interpretation of over-education, and so future research should also examine whether such human capital depreciation means some individuals are incorrectly identified as over-educated.

Finally, questions regarding how governments should target over-education and the extent to which there would be welfare gains from preventing and resolving it are critically important for the study of over-education. For instance, how could the Australian government instigate a re-allocation of over-educated individuals to well-matched jobs? How could it reduce the incidence of over-education in Australia, and would this be achieved via a dynamic process that results in changes occurring slowly over consecutive cohorts? And could the cost of such interventions exceed the welfare gains that would result from the greater utilisation of available human capital? Future research, therefore, should consider the means by which government policy interventions could prevent and resolve instances of over-education, along with whether the re-allocation of over-educated individuals would displace already well-matched individuals and what the costs and benefits of such interventions would be.

References

Akay, A. (2012) 'Finite-Sample Comparison of Alternative Methods for Estimating Dynamic Panel Data Models', *Journal of Applied Econometrics*, 27, 1189-1204.

Alba-Ramírez, A. and M. Blázquez (2003) 'Types of Job Match, Overeducation and Labour Mobility', in F. Büchel, A. De Grip and A. Mertens (eds.) *Overeducation in Europe: Current Issues in Theory and Policy*, 65-90, Cheltenham: Edward Elgar.

Albrecht, J., P. Edin, M. Sundström and S. Vroman (1999) 'Career Interruptions and Subsequent Earnings: A Reexamination Using Swedish Data', *Journal of Human Resources*, 34, 294-311.

Allen, J. and E. de Weert (2007) 'What Do Educational Mismatches Tell Us About Skill Mismatches? A Cross-Country Analysis', *European Journal of Education*, 42, 59-73.

Allen, J. and R. van der Velden (2001) 'Educational Mismatches versus Skill Mismatches: Effects on Wages, Job Satisfaction, and On-the-job Search', *Oxford Economic Papers*, 3, 434-452.

Altonji, J. and R. Blank (1999) 'Race and Gender in the Labor Market', in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3, 3143-3259, Amsterdam: Elsevier.

Amador, L., A. Nicolás and L. Vila (2008) *The Consequences on Job Satisfaction of Job-Worker Educational and Skill Mismatches in the Spanish Labour Market: A Panel Analysis*, Working Paper 2008-32, Fundación de Estudios de Economía Aplicada (FEDEA).

Angrist, J. and J. Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, New Jersey: Princeton University Press.

Arrazola, M. and J. De Hevia (2004) 'More on the Estimation of the Human Capital Depreciation Rate', *Applied Economic Letters*, 11, 145-148.

Arulampalam, W. (1999) 'A Note on Estimated Coefficients in Random Effects Probit Models', *Oxford Bulletin of Economics and Statistics*, 61, 597-602.

Arulampalam, W., A. Booth and M. Taylor (2000) 'Unemployment Persistence', *Oxford Economic Papers*, 52, 24-50.

Arulampalam, W. and M. Stewart (2009) 'Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators', *Oxford Bulletin of Economics and Statistics*, 71, 659-681.

Australian Bureau of Statistics (ABS) (1994) *Australian Social Trends,* 1994, Cat. no. 4102.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (1997) *ASCO – Australian Standard Classification of Occupations, Second Edition*, Cat. no. 1220.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2001) *Australian Standard Classification of Education (ASCED)*, Cat. no. 1272.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2002) *Education Variables*, Cat. no. 1246.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2006) *ANZSCO – Australian and New Zealand Standard Classification of Occupations, First Edition*, Cat. no. 1220.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2007a) *Information Paper: Labour Force Survey Sample Design, November 2007, Third Edition*, Cat. no. 6269.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2007b) *Labour Statistics: Concepts, Sources and Methods, April 2007*, Cat. no. 6102.0.55.001, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2009) *Technical Manual: Census of Population and Housing: Census Sample File, Technical Paper, Australia*, Cat. no. 2037.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2010) *Education and Training Experience, Australia, 2009*, Cat. no. 6278.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2011) *Government Finance Statistics, Australia, 2009-10*, Cat. no. 5512.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) (2012) *Education and Work, Australia, 2012*, Cat. no. 6227.0, ABS, Canberra.

Australian Bureau of Statistics (ABS) *Census Tables*, Cat. no. 2068.0, various years, ABS, Canberra.

Australian Bureau of Statistics (ABS) *Education and Training Experience, Australia*, Cat. no. 6278.0, various years, ABS, Canberra.

Australian Bureau of Statistics (ABS) *Education and Work, Australia*, Cat. no. 6227.0, various years, ABS, Canberra.

Australian Bureau of Statistics (ABS) *Labour Force, Australia, Detailed, Quarterly*, Cat. no. 6291.0.55.003, various years, ABS, Canberra.

Australian Qualifications Framework (AQF) Advisory Board (2007) *Australian Qualifications Framework: Implementation Handbook,* AQF Advisory Board to MCEETYA, Victoria.

Baetschmann, G., K. Staub and R. Winkelmann (2011) *Consistent Estimation of the Fixed Effects Ordered Logit Model*, IZA Discussion Paper Series No. 5443, The Institute for the Study of Labor (IZA).

Baltagi, B. (2001) *Econometric Analysis of Panel Data*, New Jersey: John Wiley & Sons.

Battu, H., C. Belfield and P. Sloane (1999) 'Overeducation Among Graduates: A Cohort View', *Education Economics*, 7, 21-38.

Battu, H., C. Belfield and P. Sloane (2000) 'How Well Can We Measure Graduate Over-education and Its Effects?', *National Institute Economic Review*, 171, 82-93.

Bauer, T. (2002) 'Educational Mismatch and Wages: A Panel Analysis', *Economics of Education Review*, 21, 221-229.

Becker, G. (1964) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, National Bureau of Economic Research (NBER), New York: Columbia University Press.

Becker, G. (1995) *The Essence of Becker*, Stanford: Hoover Institution Press.

Bender, K. (1998) 'The Central Government-Private Sector Wage Differential', *Journal of Economic Surveys*, 12, 177-220.

Bender, K. and J. Heywood (2006) *Educational Mismatch Among PhDs: Determinants and Consequences*, NBER Working Paper Series No. 12693, National Bureau of Economic Research.

Berg, I. (1971) *Education and Jobs: The Great Training Robbery*, Boston: Beacon.

Biddle, J. and G. Zarkin (1988) 'Worker Preferences and Market Compensation for Job Risk', *Review of Economics and Statistics*, 70, 660-667.

Black, D. and J. Smith (2004) 'How Robust is the Evidence on the Effects of College Quality? Evidence from Matching', *Journal of Econometrics*, 121, 99-124.

Blinder, A. (1973) 'Wage Discrimination: Reduced Form and Structural Estimates', *Journal of Human Resources*, 8, 436-455.

Blundell, R. and M. Costa Dias (2000) 'Evaluation Methods for Non-Experimental Data', *Fiscal Studies*, 21, 427-468.

Blundell, R., L. Dearden and B. Sianesi (2004) *Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS*, IFS Working Paper W03/20, The Institute for Fiscal Studies (IFS).

Booth, A. and J. van Ours (2008) 'Job Satisfaction and Family Happiness: The Part-Time Work Puzzle', *The Economic Journal*, 118, F77-F99.

Borghans and A. De Grip (eds.) (2000) *The Overeducated Worker? The Economics of Skill Utilisation*, Cheltenham: Edward Elgar.

Borland, J., P. Dawkins, D. Johnson and R. Williams (2000) *Returns to Investment in Higher Education*, Melbourne Economics of Higher Education Research Program Research Report No. 1, Melbourne Institute of Applied Economic and Social Research.

Borland, J. and Y. Tseng (2007) 'Does a Minimum Job Search Requirement Reduce Time on Unemployment Payments? Evidence from the Jobseeker Diary in Australia', *Industrial and Labor Relations Review*, 60, 357-378.

Borland, J., Y. Tseng and R. Wilkins (2005) *Experimental and Quasi-Experimental Methods of Microeconomic Program and Policy Evaluation*, Melbourne Institute Working Paper Series No. 8/05, The University of Melbourne.

Bowles, S., H. Gintis and M. Osborne (2001) 'The Determinants of Earnings: A Behavioural Approach', *Journal of Economic Literature*, 39, 1137-1176.

Brown, C. (1980) 'Equalizing Differences in the Labour Market', *Quarterly Journal of Economics*, 94, 113-134.

Büchel, F. (2002) 'The Effects of Overeducation on Productivity in Germany – The Firms' Viewpoint', *Economics of Education Review*, 21, 263-275.

Büchel, F., A. De Grip and A. Mertens (eds.) (2003) *Overeducation in Europe: Current Issues in Theory and Policy*, Cheltenham: Edward Elgar.

Büchel, F. and A. Mertens (2004) 'Overeducation, Undereducation, and the Theory of Career Mobility', *Applied Economics*, 36, 803-816.

Buddelmeyer, H., W. Lee and M. Wooden (2010) 'Low-Paid Employment and Unemployment Dynamics in Australia', *The Economic Record*, 86, 28-48.

Cahuc, P. and A. Zylberberg (2004) *Labor Economics*, Cambridge: MIT Press.

Cameron, A. and P. Trivedi (2005) *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

Cameron, A. and P. Trivedi (2009) *Microeconometrics Using Stata*, Texas: Stata Press.

Card, D. (1999) 'The Causal Effect of Education on Earnings', in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3, 1801-1863, Amsterdam: Elsevier.

Card, D. (2001) 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems', *Econometrica*, 69, 1127-1160.

Card, D. and D. Sullivan (1988) 'Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment', *Econometrica*, 56, 497-530.

Chamberlain, G. (1980) 'Analysis of Covariance with Qualitative Data', *The Review of Economic Studies*, 47, 225-238.

Chamberlain, G. (1984) 'Panel Data', in Z. Griliches and M. Intriligator (eds.) *Handbook of Econometrics,* Volume II, Amsterdam: Elsevier Science Publishers.

Chevalier, A. (2003) 'Measuring Over-education', *Economica*, 70, 509-531.

Chevalier, A. and J. Lindley (2006) *Over-Education and the Skills of UK Graduates*, IZA Discussion Paper Series No. 2442, The Institute for the Study of Labor (IZA).

Chiswick, B. and P. Miller (2010) 'The Effects of Educational-Occupational Mismatch on Immigrant Earnings in Australia, with International Comparisons', *International Migration Review*, 44, 869-898.

Clark, K. and N. Kanellopoulos (2009) *Low Pay Persistence in European Countries*, IZA Discussion Paper Series No. 4183, The Institute for the Study of Labor (IZA).

COAG (Council of Australian Governments) (2008a) *National Agreement for Skills and Workforce Development*, Canberra, viewed 1 June 2011 (http://www.federalfinancialrelations.gov.au/content/national_agreements/downloads/IGA_FFR_ScheduleF_NSWD_National_Agreement.pdf).

COAG (Council of Australian Governments) (2008b) *National Education Agreement*, Canberra, viewed 1 June 2011 (http://www.federalfinancialrelations.gov.au/content/national_agreements/downloads/IGA_FFR_ScheduleF_national_education_agreement.pdf).

COAG (Council of Australian Governments) Reform Council (2010) *National Education Agreement: Performance Report for 2009*, COAG Reform Council, Sydney.

Cobb-Clark, D. and T. Crossley (2003) 'Econometrics for Evaluations: An Introduction to Recent Developments', *The Economic Record*, 79, 491-511.

Cohn, E. and Y. Ng (2000) 'Incidence and Wage Effects of Overschooling and Underschooling in Hong Kong', *Economics of Education Review*, 19, 159-168.

D'Addio, A., T. Eriksson and P. Frijters (2007) 'An Analysis of the Determinants of Job Satisfaction When Individuals' Baseline Satisfaction Levels May Differ', *Applied Economics*, 39, 2413-2423.

Daly, M., F. Büchel and G. Duncan (2000) 'Premiums and Penalties for Surplus and Deficit Education: Evidence from the United States and Germany', *Economics of Education Review*, 19, 169-178.

Das, M. and A. van Soest (1999) 'A Panel Data Model for Subjective Information on Household Income Growth', *Journal of Economic Behavior & Organization*, 40, 409-426.

Dearden, L., H. Reed and J. Van Reenen (2006) 'The Impact of Training on Productivity and Wages: Evidence from British Panel Data', *Oxford Bulletin of Economics and Statistics*, 68, 397-421.

De Grip, A. (2006) *Evaluating Human Capital Obsolescence*, Working Paper, Research Centre for Education and the Labour Market, Faculty of Economics and Business Administration, Maastricht University.

De Grip, A., H. Bosma, D. Willems, and M. van Boxtel (2008) 'Job-Worker Mismatch and Cognitive Decline', *Oxford Economic Papers*, 60, 237-253.

De Grip, A. and J. Van Loo (2002) 'The Economics of Skill Obsolescence: A Review', *The Economics of Skill Obsolescence*, 21, 1-26.

Dehejia, R. and S. Wahba (2002) 'Propensity Score Matching for Nonexperimental Causal Studies', *Review of Economics and Statistics*, 84, 151-161.

Dekker, R., A. De Grip and H. Heijke (2002) 'The Effects of Training and Overeducation on Career Mobility in a Segmented Labour Market', *International Journal of Manpower*, 23, 106-125.

Di Pietro, G. and P. Urwin (2006) 'Education and Skills Mismatch in the Italian Graduate Labour Market', *Applied Economics*, 38, 79-93.

Doiron, D. (2004) 'Welfare Reform and the Labour Supply of Lone Parents in Australia: A Natural Experiment Approach', *The Economic Record*, 80, 157-176.

Dolton, P. and M. Silles (2003) 'The Determinants and Consequences of Graduate Overeducation', in F. Büchel, A. De Grip and A. Mertens (eds.) *Overeducation in Europe: Current Issues in Theory and Policy*, 189-216, Cheltenham: Edward Elgar.

Dolton, P. and M. Silles (2008) 'The Effects of Over-education on Earnings in the Graduate Labour Market', *Economics of Education Review*, 27, 125-139.

Dolton, P. and A. Vignoles (2000) 'The Incidence and Effects of Overeducation in the UK Graduate Labour Market', *Economics of Education Review*, 19, 179-198.

Dorn, D. and A. Sousa-Poza (2006) *Overqualification: Permanent or Transitory?*, Working Paper, University of St. Gallen.

Duncan, G. and S. Hoffman (1981) 'The Incidence and Wage Effects of Overeducation', *Economics of Education Review*, 1, 75-86.

Ehrenberg, R. and R. Smith (2006) *Modern Labor Economics*, Illinois: Pearson Education.

Elias, P. and K. Purcell (2004) 'Is Mass Higher Education Working? Evidence from the Labour Market Experiences of Recent Graduates', *National Institute Economic Review*, 190, 60-74.

European Centre for the Development of Vocational Training (CEDEFOP) (2009) *Skill Mismatch: Identifying Priorities for Future Research*, CEDEFOP Working Paper No. 3.

Felstead, A., D. Gallie, F. Green and Y. Zhou (2007) *Skills at Work, 1986 to 2006*, ESRC Centre on Skills, Knowledge and Organisational Performance, Universities of Oxford and Cardiff, Oxford.

Ferrer-i-Carbonell, A. and P. Frijters (2004) 'How Important is Methodology for the Estimates of the Determinants of Happiness?', *The Economic Journal*, 114, 641-659.

Fitzgerald, J., P. Gottschalk and R. Moffitt (1998) 'An Analysis of Sample Attrition in Panel Data', *Journal of Human Resources*, 33, 251-299.

Fleming, C. and P. Kler (2008) 'I'm Too Clever For This Job: A Bivariate Probit Analysis on Overeducation and Job Satisfaction in Australia', *Applied Economics*, 40, 1123-1138.

Frank, R. (1978) 'Why Women Earn Less: The Theory and Estimation of Differential Overqualification', *The American Economic Review*, 68, 360-373.

Freeman, R. (1976) *The Overeducated American*, New York: Academic Press.

Frenette, M. (2004) 'The Overqualified Canadian Graduate: The Role of the Academic Program in the Incidence, Persistence, and Economic Returns to Overqualification', *Economics of Education Review*, 23, 29-45.

Frijters, P., J. Haisken-DeNew and M. Shields (2004a) 'Investigating the Patterns and Determinants of Life Satisfaction in Germany Following Reunification', *Journal of Human Resources*, 39, 649-674.

Frijters, P., J. Haisken-DeNew and M. Shields (2004b) 'Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification', *The American Economic Review*, 94, 730-740.

Girma, S. and H. Görg (2007) 'Evaluating the Foreign Ownership Wage Premium Using a Difference-in-differences Matching Approach', *Journal of International Economics*, 72, 97-112.

Görlich, D. and A. De Grip (2009) 'Human Capital Depreciation During Hometime', *Oxford Economic Papers*, 61, 98-121.

Granger, C. (2007) 'Causality in Economics', in P. Machamer and G. Wolters (eds.) *Thinking About Causes: From Greek Philosophy to Modern Physics*, 284-296, Pittsburgh: University of Pittsburgh Press.

Green, C., P. Kler and G. Leeves (2007) 'Immigrant Overeducation: Evidence from Recent Arrivals to Australia', *Economics of Education Review*, 26, 420-432.

Green, F. and S. McIntosh (2007) 'Is There a Genuine Under-utilization of Skills Amongst the Over-qualified?', *Applied Economics*, 39, 427-439.

Green, F., S. McIntosh and A. Vignoles (1999) *'Overeducation' and Skills – Clarifying the Concepts*, London: Centre for Economic Performance.

Green, F. and Y. Zhu (2010) 'Overqualification, Job Dissatisfaction, and Increasing Dispersion in the Returns to Graduate Education', *Oxford Economic Papers*, 62, 740-763.

Greene, W. (2004) 'The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects', *Econometrics Journal*, 7, 98-119.

Groeneveld, S. and J. Hartog (2004) 'Overeducation, Wages and Promotions Within the Firm', *Labour Economics*, 11, 701-714.

Groot, L. and A. Hoek (2000) 'Job Competition in the Dutch Labour Market', in L. Borghans and A. De Grip (eds.) *The Overeducated Worker? The Economics of Skill Utilisation*, 231-252, Cheltenham: Edward Elgar.

Groot, W. and H. Maassen van den Brink (2000a) 'Overeducation in the Labor Market: A Meta-Analysis', *Economics of Education Review*, 19, 149-158.

Groot, W. and H. Maassen van den Brink (2000b) 'Skill Mismatches in the Dutch Labor Market', *International Journal of Manpower*, 21, 584-595.

Groot, W. and H. Maassen van den Brink (2003) 'The Dynamics of Skill Mismatches in the Dutch Labour Market', in F. Büchel, A. De Grip and A. Mertens (eds.) *Overeducation in Europe: Current Issues in Theory and Policy*, 49-64, Cheltenham: Edward Elgar.

Gruetter, M. and R. Lalive (2009) 'The Importance of Firms in Wage Determination', *Labour Economics*, 16, 149-160.

Guironnet, J.-P. and N. Peypoch (2007) 'Human Capital Allocation and Overeducation: A Measure of French Productivity (1987, 1999)', *Economic Modelling*, 24, 398-410.

Halaby, C. (1994) 'Overeducation and Skill Mismatch', *Sociology of Education*, 67, 47-59.

Hartog, J. (1997) 'Wandering Along the Hills of ORU Land', in J. Heijke (ed.) *Education, Training and Employment in the Knowledge Based Economy*, London: Macmillan.

Hartog, J. (2000) 'Over-education and Earnings: Where Are We, Where Should We Go?', *Economics of Education Review*, 19, 131-147.

Hartog, J. and H. Maassen van den Brink (eds.) (2007) *Human Capital: Advances in Theory and Evidence*, Melbourne: Cambridge University Press.

Heckman, J. (1979) 'Sample Selection Bias as a Specification Error', *Econometrica*, 47, 153-161.

Heckman, J. (1981a) 'Heterogeneity and State Dependence', in S. Rosen (ed.) *Studies in Labor Markets*, 91-140, Chicago: University of Chicago Press.

Heckman, J. (1981b) 'Statistical Models for Discrete Panel Data', in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, 114-178, Cambridge: MIT Press.

Heckman, J. (1981c) 'The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process', in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, 179-195, Cambridge: MIT Press.

Heckman, J. (2008) 'Econometric Causality', *International Statistical Review*, 76, 1-27.

Heckman, J. and G. Borjas (1980) 'Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence', *Economica*, 47, 247-283.

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998) 'Characterizing Selection Bias Using Experimental Data', *Econometrica*, 66, 1017-1098.

Heckman, J., H. Ichimura and P. Todd (1997) 'Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *Review of Economic Studies*, 64, 605-654.

Heckman, J., L. Lochner and P. Todd (2006) 'Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond', in E. Hanushek and F. Welch (eds.) *Handbook of the Economics of Education*, Volume 1, 307-458, Amsterdam: Elsevier.

Heckman, J., L. Lochner and P. Todd (2008) 'Earnings Functions and Rates of Return', *Journal of Human Capital*, 2, 1-31.

Heckman, J. and R. Robb (1985) 'Alternative Methods for Evaluating the Impact of Interventions', in M. Fix and R. Struyk (eds.) *Longitudinal Analysis of Labor Market Data*, 156-246, Cambridge: Cambridge University Press.

Heckman, J. and R. Willis (1977) 'A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women', *Journal of Political Economy*, 85, 27-58.

Hersch, J. (1995) 'Optimal 'Mismatch' and Promotions', *Economic Inquiry*, 33, 611-624.

Ho, D., K. Imai, G. King and E. Stuart (2007) 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference', *Political Analysis*, 15, 199-236.

Holland, P. (1986) 'Statistics and Causal Inference', *Journal of the American Statistical Association*, 81, 945-960.

Honore, B. and E. Kyriazidou (2000) 'Panel Data Discrete Choice Models with Lagged Dependent Variables', *Econometrica*, 68, 839-874.

Hwang, H., R. Reed and C. Hubbard (1992) 'Compensating Wage Differentials and Unobserved Productivity', *Journal of Political Economy*, 100, 835-858.

Imai, K., L. Keele, D. Tingley and T. Yamamoto (2011) 'Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies', *American Political Science Review*, 105, 765-789.

Imbens, G. and J. Wooldridge (2009) 'Recent Developments in the Econometrics of Program Evaluation', *Journal of Economic Literature*, 47, 5-86.

Jensen, U. (2003) 'Measuring Overeducation with Earnings Frontiers and Panel Data', in F. Büchel, A. De Grip and A. Mertens (eds.) *Overeducation in Europe: Current Issues in Theory and Policy*, 155-169, Cheltenham: Edward Elgar.

Jeon, S., G. Kalb and H. Vu (2011) 'The Dynamics of Welfare Participation Among Women Who Experienced Teenage Motherhood in Australia', *The Economic Record*, 87, 235-251.

Jones, M., K. Mavromaras, P. Sloane and Z. Wei (2011) *Disability and Job Mismatches in the Australian Labour Market*, IZA Discussion Paper Series No. 6152, The Institute for the Study of Labor (IZA).

Jones, A. and S. Schurer (2011) 'How Does Heterogeneity Shape the Socioeconomic Gradient in Health Satisfaction?', *Journal of Applied Econometrics*, 26, 549-579.

Karmel, T. (2009) *Global Skills Crunch: A Case of Dog Eat Dog?*, NCVER Conference Paper, National Centre for Vocational Education Research (NCVER).

Kassenboehmer, S. and J. Haisken-DeNew (2009) 'You're Fired! The Causal Negative Effect of Entry Unemployment on Life Satisfaction', *The Economic Journal*, 119, 448-462.

Katz, L. and D. Autor (1999) 'Changes in the Wage Structure and Earnings Inequality', in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3, 1463-1555, Amsterdam: Elsevier.

Kiker, B., M. Santos and M. Mendes de Oliveira (1997) 'Overeducation and Undereducation: Evidence for Portugal', *Economics of Education Review*, 16, 111-125.

Kler, P. (2005) 'Graduate Overeducation in Australia: A Comparison of the Mean and Objective Methods', *Education Economics*, 13, 47-72.

Kler, P. (2006) *Graduate Overeducation in Australia*, PhD dissertation, The University of Queensland.

Kler, P. (2007) 'A Panel Data Investigation into Over-education Among Tertiary Educated Australian Immigrants', *Journal of Economic Studies*, 34, 179-193.

Koopmans, T. and M. Beckmann (1957) 'Assignment Problems and the Location of Economic Activities', *Econometrica*, 25, 90-105.

Korpi, T. and M. Tåhlin (2009) 'Educational Mismatch, Wages, and Wage Growth: Overeducation in Sweden, 1974-2000', *Labour Economics*, 16, 183-193.

Kunze, A. (2002) *The Timing of Careers and Human Capital Depreciation*, IZA Discussion Paper Series No. 509, The Institute for the Study of Labor (IZA).

Lamo, A. and J. Messina (2010) 'Formal Education, Mismatch and Wages After Transition: Assessing the Impact of Unobserved Heterogeneity Using Matching Estimators', *Economics of Education Review*, 29, 1086-1099.

Lancaster, T. (2000) 'The Incidental Parameter Problem Since 1948', *Journal of Econometrics*, 95, 391-413.

Lee. W. and M. Coelli (2010) 'The Labour Market Effects of Vocational Education and Training in Australia', *The Australian Economic Review*, 43, 389-408.

Lee, W. and U. Oguzoglu (2007) 'Income Support and Stigma Effects for Young Australians', *The Australian Economic Review*, 40, 369-384.

Leuven, E. and H. Oosterbeek (2011) 'Overeducation and Mismatch in the Labor Market', in E. Hanushek, S. Machin and L. Woessmann (eds.) *Handbook of The Economics of Education*, Volume 4, 283-326, Oxford: Elsevier.

Lindley, J. and S. McIntosh (2008) *A Panel Data Analysis of the Incidence and Impact of Overeducation*, Discussion Paper, Department of Economics, University of Sheffield.

Linsley, I. (2005) 'Causes of Overeducation in the Australian Labour Market', *Australian Journal of Labour Economics*, 8, 121-143.

Lorgelly, P. and J. Lindley (2008) 'What is the Relationship Between Income Inequality and Health? Evidence from the BHPS', *Health Economics*, 17, 249-265.

Machamer, P. and G. Wolters (eds.) (2007) *Thinking About Causes: From Greek Philosophy to Modern Physics*, Pittsburgh: University of Pittsburgh Press.

Mavromaras, K. and S. McGuinness (2007) 'Education and Skill Mismatches in the Labour Market: Editors Introduction', *The Australian Economic Review*, 40, 279-285.

Mavromaras, K. and S. McGuinness (2012) 'Overskilling Dynamics and Education Pathways', *Economics of Education Review*, 31, 619-628.

Mavromaras, K., S. McGuinness and Y. Fok (2009) 'Assessing the Incidence and Wage Effects of Overskilling in the Australian Labour Market', *The Economic Record*, 85, 60-72.

Mavromaras, K., S. McGuinness and Y. Fok (2010a) *The Incidence and Wage Effects of Overskilling Among Employed VET Graduates*, NCVER Monograph Series 03/2009, National Centre for Vocational Education Research (NCVER).

Mavromaras, K., S. McGuinness, N. O'Leary, P. Sloane and Y. Fok (2010b) 'The Problem of Overskilling in Australia and Britain', *The Manchester School*, 78, 219-241.

Mavromaras, K., S. McGuinness, N. O'Leary, P. Sloane and Z. Wei (2010c) *Job Mismatches and Labour Market Outcomes: Panel Evidence on Australian University Graduates*, IZA Discussion Paper Series No. 5083, The Institute for the Study of Labor (IZA).

Mavromaras, K., S. McGuinness, S. Richardson, P. Sloane and Z. Wei (2011) *Over-skilling and Job Satisfaction in the Australian Labour Force*, National Vocational Education and Training Research and Evaluation Program Report, National Centre for Vocational Education Research (NCVER).

Mavromaras, K., S. McGuinness and M. Wooden (2007) 'Overskilling in the Australian Labour Market', *The Australian Economic Review*, 40, 307-312.

McGoldrick, K. and J. Robst (1996) 'Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification', *The American Economic Review*, 86, 280-284.

McGuinness, S. (2006) 'Overeducation in the Labour Market', *Journal of Economic Surveys*, 20, 387-418.

McGuinness, S. (2008) 'How Biased are the Estimated Wage Impacts of Overeducation? A Propensity Score Matching Approach', *Applied Economics Letters*, 15, 145-149.

McGuinness, S. and J. Bennett (2007) 'Overeducation in the Graduate Labour Market: A Quantile Regression Approach', *Economics of Education Review*, 26, 521-531.

McGuinness, S. and P. Sloane (2011) 'Labour Market Mismatch Among UK Graduates: An Analysis Using REFLEX Data', *Economics of Education Review*, 30, 130-145.

McGuinness, S. and M. Wooden (2009) 'Overskilling, Job Insecurity, and Career Mobility', *Industrial Relations*, 48, 265-286.

McMillen, D., P. Seaman and L. Singell (2007) 'A Mismatch Made in Heaven: A Hedonic Analysis of Overeducation and Undereducation', *Southern Economic Journal*, 73, 901-930.

Mendes de Oliveira, M., M. Santos and B. Kiker (2000) 'The Role of Human Capital and Technological Change in Overeducation', *Economics of Education Review*, 19, 199-206.

Messinis, G. (2008) *Overeducation and Overskilling: Second Generation Australians*, Centre for Strategic Economic Studies Working Paper No. 41, Victoria University.

Messinis, G. and N. Olekalns (2007a) *Skill Mismatch and Returns to Training in Australia: Some New Evidence*, Department of Economics Research Paper No. 997, The University of Melbourne.

Messinis, G. and N. Olekalns (2007b) 'Skill Mismatch and Training in Australia: Some Implications for Policy', *The Australian Economic Review*, 40, 300-306.

Miller, P. (2007) 'Overeducation and Undereducation in Australia', *The Australian Economic Review*, 40, 292-299.

Mincer, J. (1958) 'Investment in Human Capital and Personal Income Distribution', *Journal of Political Economy*, 66, 281-302.

Mincer, J. (1974) *Schooling, Experience and Earnings*, New York: Columbia University Press.

Mincer, J. and H. Ofek (1982) 'Interrupted Work Careers: Depreciation and Restoration of Human Capital', *Journal of Human Resources*, 17, 3-24.

Mincer, J. and S. Polachek (1974) 'Family Investments in Human Capital: Earnings of Women', *Journal of Political Economy*, 82, 76-108.

Morgan, S. and C. Winship (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge: Cambridge University Press.

Mukherjee, B., J. Ahn, I. Liu, P. Rathouz and B. Sanchez (2008) 'Fitting Stratified Proportional Odds Models by Amalgamating Conditional Likelihoods', *Statistics in Medicine*, 27, 4950-4971.

Mundlak, Y. (1978) 'On the Pooling of Time Series and Cross Section Data', *Econometrica*, 46, 69-85.

Neal, D. and S. Rosen (2000) 'Theories of the Distribution of Earnings', in A. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution*, Volume 1, 379-427, Amsterdam: Elsevier.

Norris, K., R. Kelly and M. Giles (2005) *Economics of Australian Labour Markets*, New South Wales: Pearson Education Australia.

Oaxaca, R. (1973) 'Male-Female Wage Differentials in Urban Labor Markets', *International Economic Review*, 14, 693-709.

Oguzoglu, U. (2011) 'Severity of Work Disability and Work', *The Economic Record*, 87, 370-383.

Oosterbeek, H. (2000) 'Introduction to Special Issue on Overschooling', *Economics of Education Review*, 19, 129-130.

Orme, C. (1997) *The Initial Conditions Problem and Two-Step Estimation in Discrete Panel Data Models*, mimeo, Department of Economics, University of Manchester.

Ortiz, L. (2010) 'Not the Right Job, But a Secure One: Over-education and Temporary Employment in France, Italy and Spain', *Work, Employment and Society*, 24, 47-64.

Pearl, J. (2000) *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Piracha, M., M. Tani and F. Vadean (2010) *Immigrant Over- and Under-education: The Role of Home Country Labour Market Experience*, IZA Discussion Paper Series No. 5302, The Institute for the Study of Labor (IZA).

Pissarides, C. (1992) 'Loss of Skill During Unemployment and the Persistence of Employment Shocks', *Quarterly Journal of Economics*, 107, 1371-1391.

Pollmann-Schult, M. and F. Büchel (2004) 'Career Prospects of Overeducated Workers in West Germany', *European Sociological Review*, 20, 321-331.

Preston, A. (1997) 'Where Are We Now With Human Capital Theory in Australia?', *The Economic Record*, 73, 51-78.

Quinn, M. and S. Rubb (2006) 'Mexico's Labor Market: The Importance of Education-Occupation Matching on Wages and Productivity in Developing Countries', *Economics of Education Review*, 25, 147-156.

Robst, J. (1995) 'College Quality and Overeducation', *Economics of Education Review*, 14, 221-228.

Robst, J. (2007) 'Education and Job Match: The Relatedness of College Major and Work', *Economics of Education Review*, 26, 397-407.

Rosen, S. (1986) 'The Theory of Equalizing Differences', in O. Ashenfelter and R. Layard (eds.) *Handbook of Labor Economics*, Volume I, 641-692, Amsterdam: Elsevier.

Rosenbaum, P. and D. Rubin (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, 70, 41-55.

Roy, A. (1951) 'Some Thoughts on the Distribution of Earnings', *Oxford Economic Papers*, 3, 135-146.

Rubb, S. (2003a) 'Overeducation: A Short or Long Run Phenomenon for Individuals?', *Economics of Education Review*, 22, 389-394.

Rubb, S. (2003b) 'Overeducation in the Labour Market: A Comment and Re-Analysis of a Meta-Analysis', *Economics of Education Review*, 22, 621-629.

Rubb, S. (2006) 'Educational Mismatches and Earnings: Extensions of Occupational Mobility Theory and Evidence of Human Capital Depreciation', *Education Economics*, 14, 135-154.

Rubin, D. (1974) 'Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies', *Journal of Educational Psychology*, 66, 688-701.

Rubin, D. and N. Thomas (2000) 'Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates', *Journal of the American Statistical Association*, 95, 573-585.

Rumberger, R. (1987) 'The Impact of Surplus Schooling on Productivity and Earnings', *Journal of Human Resources*, 22, 24-50.

Ryan, C. and M. Sinning (2009) *Skill Matches to Job Requirements*, National Vocational Education and Training Research and Evaluation Program Report, National Centre for Vocational Education Research (NCVER).

Ryan, C. and M. Sinning (2011a) *Differing Skill Requirements Across Countries and Over Time*, National Vocational Education and Training Research and Evaluation Program Report, National Centre for Vocational Education Research (NCVER).

Ryan, C. and M. Sinning (2011b) *Skill (Mis-)Matches and Over-education of Younger Workers*, National Vocational Education and Training Research and Evaluation Program Report, National Centre for Vocational Education Research (NCVER).

Sattinger, M. (1993) 'Assignment Models of the Distribution of Earnings', *Journal of Economic Literature*, 31, 831-880.

Schmitz, H. (2011) 'Why are the Unemployed in Worse Health? The Causal Effect of Unemployment on Health', *Labour Economics*, 18, 71-78.

Schultz, T. (1961) 'Investment in Human Capital', *The American Economic Review*, 51, 1-17.

Sicherman, N. (1991) "Overeducation' in the Labour Market', *Journal of Labour Economics*, 9, 101-122.

Sicherman, N. and O. Galor (1990) 'A Theory of Career Mobility', *Journal of Political Economy*, 98, 169-192.

Sloane, P. (2003) 'Much Ado About Nothing? What Does the Overeducation Literature Really Tell Us?', in F. Büchel, A. De Grip and A. Mertens (eds.) *Overeducation in Europe: Current Issues in Theory and Policy*, 11-45, Cheltenham: Edward Elgar.

Sloane, P. (2007) 'Overeducation in the United Kingdom', *The Australian Economic Review*, 40, 286-291.

Sloane, P., H. Battu and P. Seaman (1999) 'Overeducation, Undereducation and the British Labour Market', *Applied Economics*, 31, 1437-1453.

Smith, J. and P. Todd (2005) 'Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?', *Journal of Econometrics*, 125, 305-353.

Spence, M. (1973) 'Job Market Signalling', *Quarterly Journal of Economics*, 87, 355-374.

Stewart, M. (2007) 'The Interrelated Dynamics of Unemployment and Low-Wage Employment', *Journal of Applied Econometrics*, 22, 511-531.

Stewart, M. and J. Swaffield (1999) 'Low Pay Dynamics and Transition Probabilities', *Economica*, 66, 23-42.

Summerfield, M. (ed.) (2010) *HILDA User Manual – Release 9*, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Thurow, L. (1975) *Generating Inequality*, New York: Basic Books.

Tinbergen, J. (1951) 'Some Remarks on the Distribution of Labour Incomes', *International Economic Papers, Translations Prepared for the International Economic Association*, 1, 195-207.

Tinbergen, J. (1956) 'On the Theory of Income Distribution', *Weltwirtschaftliches Archiv*, 77, 155-175.

Tsai, Y. (2010) 'Returns to Overeducation: A Longitudinal Analysis of the US Labor Market', *Economics of Education Review*, 29, 606-617.

Tsang, M. (1987) 'The Impact of Underutilization of Education on Productivity: A Case Study of the US Bell Companies', *Economics of Education Review*, 6, 239-254.

Tsang, M., R. Rumberger and H. Levin (1991) 'The Impact of Surplus Schooling on Worker Productivity', *Industrial Relations*, 30, 209-228.

Vahey, S. (2000) 'The Great Canadian Training Robbery: Evidence on the Returns to Educational Mismatch', *Economics of Education Review*, 19, 219-227.

Van der Meer, P. (2006) 'The Validity of Two Education Requirement Measures', *Economics of Education Review*, 25, 211-219.

Verdugo, R. and N. Verdugo (1992) 'Surplus Schooling and Earnings: Reply to Cohn and to Gill and Sollberg', *Journal of Human Resources*, 22, 690-695.

Verhaest, D. and E. Omey (2006a) 'Discriminating Between Alternative Measures of Over-education', *Applied Economics*, 38, 2113-2120.

Verhaest, D. and E. Omey (2006b) 'Measuring the Incidence of Over- and Undereducation', *Quality & Quantity*, 40, 783-803.

Verhaest, D. and E. Omey (2006c) 'The Impact of Overeducation and its Measurement', *Social Indicators Research*, 77, 419-448.

Verhaest, D. and E. Omey (2009) *Overeducation and Earnings: Some Further Panel-Data Evidence*, HUB Research Paper 2009/08, HUB Research Papers in Economics and Management.

Verhaest, D. and T. Schatteman (2010) *Overeducation in the Early Career: An Analysis Using Sequence Techniques*, HUB Research Paper 2010/09, HUB Research Papers in Economics and Management.

Voon, D. and P. Miller (2005) 'Undereducation and Overeducation in the Australian Labour Market', *The Economic Record*, 81, S22-S33.

Vu, H. (2010) *The Dynamics of Income Support in Australia*, PhD dissertation, The University of Melbourne.

Wald, S. (2005) 'The Impact of Overqualification on Job Search', *International Journal of Manpower*, 26, 140-156.

Watson, N. (2004) *Wave 2 Weighting*, HILDA Project Technical Paper Series, No. 4/04, The University of Melbourne.

Watson, N. and M. Summerfield (2009) *Quality of the Occupation and Industry Coding in the HILDA Survey*, HILDA Project Discussion Paper Series No. 3/09, The University of Melbourne.

Watson, N. and M. Wooden (2004) 'Sample Attrition in the HILDA Survey', *Australian Journal of Labour Economics*, 7, 293-308.

Watson, N. and M. Wooden (2009) 'Identifying Factors Affecting Longitudinal Survey Response', in P. Lynn (ed.) *Methodology of Longitudinal Surveys*, 157-179, Chichester: John Wiley and Sons.

Watson, N. and M. Wooden (2011) *Re-engaging with Survey Non-respondents: The BHPS, SOEP and HILDA Survey Experience*, HILDA Project Discussion Paper Series No. 1/11, The University of Melbourne.

Wooden, M. and N. Watson (2007) 'The HILDA Survey and its Contribution to Economic and Social Research (So Far)', *The Economic Record*, 83, 208-231.

Wooldridge, J. (2003) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Wooldridge, J. (2005) 'Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity', *Journal of Applied Econometrics*, 20, 39-54.

Appendices

Appendix 1.1: Key definitions, assumptions and limitations of study

Definition: *Over-education*

> An individual who has human capital from education that is under-utilised in their current job is *over-educated*. [D1]

Theoretical assumption: Existence of over-education

For over-education to occur in labour markets, it must be assumed that:

> Jobs have *productivity ceilings*—thresholds beyond which output becomes unresponsive to the human capital possessed by individuals—that can prevent individuals from fully utilising their human capital in performing the job. [A1]

That is, for each job, there is a particular level of human capital—referred to as the *required human capital level*—at which additional human capital ceases to increase the output produced (or labour productivity achieved) by individuals.[1] As defined in [D1], comparing the human capital individuals derive from education with these required human capital levels then determines whether individuals are over-educated for their current job.

Empirical assumptions: Identification of over-education

To empirically identify instances of over-education, it is assumed that:

> Levels of education (or qualifications)—measured in this study using AQF qualifications—can be used to quantify the required human capital levels of jobs; thus, a *required education level* is defined for each job. [A2]

> Individuals who complete the same level of education (or AQF qualification) acquire the same level of human capital from it. [A3]

Given [A2] and [A3], over-education is empirically identified using information on the level of education completed by each individual and the required education level of their current job: an individual is deemed *over-educated* if their education (or highest AQF qualification) level exceeds that considered necessary to perform their job.

[1] In reality, the relationship between human capital and output may be better characterised as a continuous production function with diminishing returns to human capital and, as a result, required human capital levels interpreted as either a form of *optimal* human capital level or the point at which only a *fraction* of additional human capital increases output. Additionally, required human capital levels are with respect to *relevant* human capital, where relevance refers to subject matter (e.g., civil engineering, nursing and accounting). For the discussion of these issues see Section 2.2.

Limitations

Based on the empirical identification of over-education, this study has the following limitations:

Given [A2] and [A3], the study does not consider the potential for individuals' innate abilities, quality of educational institution attended and qualification vintage (or year completed) to affect the human capital derived from education. [L1]

Given the use of each individual's highest AQF qualification, the study does not consider whether the completion of multiple qualifications—at the same ASCED level or different levels—may render an individual over-educated. [L2]

The study does not consider whether the human capital individuals derive from education is entirely relevant for their job, in terms of subject matter (where subject matter refers to the field of study (or, as defined in the ASCED, the *field of education*) of individuals' qualifications). [L3]

Appendix 2.1: Alternative theoretical frameworks for over-education

Assignment models

Assignment models, as outlined by Sattinger (1993), are based on the assumption that economies are faced with an allocation problem: individuals of varying levels of ability must be allocated to jobs of varying levels of complexity.[1] These models implicitly assume that jobs can impose productivity ceilings on individuals. Ultimately, the quality of these individual-job matches affects the labour productivity achieved by individuals, and hence their wage, and it affects the total output of the economy. Numerous models have been developed to explain the solution of this allocation problem, such as the models of Tinbergen (1951), Roy (1951) and Koopmans and Beckmann (1957). Despite their differences, these models all share a similar structure and it is considered the general set-up for assignment models. Essentially, the models specify the following for an economy: the set of jobs available; the differences that exist between individuals (in terms of ability); the technology that translates the characteristics of individuals and jobs into output; and, the mechanism that assigns individuals to jobs (Sattinger, 1993; McGuinness, 2006).

An important element of the assignment mechanism is that individuals are free to choose jobs from any sector of the economy, such that their earnings are not dependent on their performance in any one sector and poor performance in one sector can lead them to employment in another sector. Thus, assignment is a result of individuals' choices (i.e., self-selection). This choice of jobs and sectors also introduces an intermediate step between individuals' characteristics and their earnings, which means earnings functions are no longer directly observable relationships (Sattinger, 1993; McGuinness, 2006). Instead, they are the direct result of the particular equilibrium that solved the allocation problem, whereby the wage differentials arise from the establishing of equality in labour supplied and demanded in each of the economy's labour markets (Sattinger, 1993).[2] The higher wages paid to individuals with certain characteristics, therefore, aid the allocation of individuals across jobs, rather than merely serving as rewards for possessing such characteristics (McGuinness, 2006). As a result, earnings functions are interpreted as hedonic wage functions that represent reduced-form rather than structural relationships (Sattinger, 1993).

---

[1] In assignment models, individuals' ability levels are considered to be their productive capacity (or potential labour productivity levels) in various jobs, and it is typically assumed individuals may have comparative advantages in certain jobs (Sattinger, 1993). Also, education enhances such ability levels (as do the other elements of human capital).

[2] Thus, any changes in the supply of either individuals or jobs will ultimately lead to a new equilibrium; one with a different earnings function and, therefore, a different earnings distribution. According to Sattinger (1993), this ability to explain changes in earnings distributions was a prime reason for the development of assignment models as previously, based on human capital theory of labour markets, it was difficult to explain such changes (which are observed in empirical research).

In assignment models, over-education occurs when individual-job matches lead to productivity ceilings being imposed on individuals, where (some of) the subsequently unused abilities were derived from education. For example, an assignment mechanism could entail individuals and jobs being sorted top-down by ability and complexity level and then matched: the most able individual assigned the most complex job and so on until the least able individual is assigned the least complex job (Allen and van der Velden, 2001). In such a scenario, over-education may arise in equilibrium if there are differences in the distributions of individuals by ability and jobs by complexity (e.g., more highly able individuals than highly complex jobs in the economy). Such instances of over-education would persist until a new equilibrium emerges (i.e., individuals may be persistently over-educated).[3]

Job signalling model

In the job signalling model of Spence (1973), firms seek to pay individuals a wage equal to the value of their labour productivity in a particular job, but they are unable to perfectly observe individuals' productive capacity prior to hiring them. It is implicitly assumed that such imperfect information may result in productivity ceilings being imposed on individuals. To distinguish between individuals of different productive capacities (and determine job offers), it is assumed firms use signals; these are costly for individuals to acquire, with costs negatively correlated with their productive capacity.[4] Education is one such signal.[5] As a result, the model assumes the primary role of education is to signal individuals' productive capacity in labour markets. An important assumption, particularly for the study of over-education, is whether this is education's only role. In the example presented by Spence (1973), it was indeed assumed education has no effect on individuals' productive capacity, but this assumption can be relaxed. The key is that firms use education to form expectations regarding the productivity of potential workers and then, based on these expectations, make job (and wage) offers. Meanwhile, individuals know the hiring decisions of firms are made using education as a signal, and so they invest in education to increase their chances of obtaining high wage jobs.[6]

[3] A key implication of assignment models is that to explain earnings differences across individuals it is necessary to consider both individual and job characteristics; this has been embraced in empirical research, including the over-education literature (McGuinness, 2006). In particular, it has been used to justify the inclusion of required education measures (a job characteristic) in earnings functions (i.e., the Duncan and Hoffman (1981) extension of Mincer earnings functions).
[4] Important underlying assumptions are that hiring individuals is costly for firms (e.g., the time and effort it takes to hire, train and supervise workers) and that firms seek to minimise such costs (i.e., maximise profits) (Spence, 1973; Kler, 2006).
[5] Education is a suitable signal because it is costly to obtain and these costs are likely lower for the more productive (or able) individuals (e.g., pecuniary costs are lower as they are more likely to obtain scholarships and complete studies more quickly, while non-pecuniary costs are lower as completion likely requires less effort). Thus, it is reasonable to assume that, on average, more able individuals will obtain more education (or higher qualifications).
[6] This assumption that individuals know firms use education as a signal can also be relaxed. It is only necessary that individuals perceive a return to education (in the form of higher wages) and that they seek to maximise the difference between their expected wages and the costs of completing different levels of education (Spence, 1973; Kler, 2006).

Ultimately, equilibrium outcomes of the job signalling model may lead to over-education. Specifically, in situations where either the costs of completing education are low or the expectations of individuals or firms regarding the completion of education are inflated, then it may be optimal for individuals to complete a level of education greater than actually required to perform their job (Linsley, 2005; Kler, 2006).[7, 8] Completing excess education is optimal for some individuals because they are seeking to protect their relative position in the distribution of potential workers (McGuinness, 2006). Since it arises in equilibrium, the job signalling model predicts such over-education will be persistent for individuals.

Job competition model

The job competition model of Thurow (1975) is based on the premise that individuals' labour productivity achieved in the workplace, and hence their wage, is determined entirely by the characteristics of their job. In particular, the model assumes that the skills used in the workplace are acquired solely from on-the-job training and, therefore, education does not affect individuals' productive capacity.[9] Labour markets are consequently markets in which individuals compete for job opportunities based on their relative training costs to firms (Thurow, 1975; Kler, 2006). Ultimately, individuals are allocated to jobs via the matching of a job queue and a labour queue; the job queue ranks jobs by their wages, while the labour queue ranks individuals by their potential training costs.[10] Since it is assumed firms use education as a signal for potential training costs (i.e., training costs should be lower for individuals with more education), the labour queue actually ranks individuals by their education level (Kler, 2006). Then, similar to assignment models, the job queue and the labour queue are matched, whereby the most highly educated (or top-ranked) individual is matched to the highest wage job and so on down both queues (Thurow, 1975).

---

[7] This outcome is often also referred to as *qualification inflation*: reduction in the costs of education leads less able individuals to complete more education, thereby weakening the effectiveness of education as a signal, which leads firms to increase the educational requirements for obtaining jobs without actually changing the jobs (Green et al., 1999). This may also be referred to as *credentialism*, though some argue the two are not always equivalent (see, for example, Berg (1971)). A further similar concept is *grade drift*, which concerns changes in educational standards over time (e.g., decline in the quality of given qualifications) and the response of firms (e.g., changing educational requirements for obtaining jobs) (Green et al., 1999). The concepts, however, are not equivalent; grade drift concerns actual changes in quality of qualifications, but qualification inflation concerns the effectiveness of qualifications as signals for individuals' productive capacity.

[8] If it is assumed that education does not enhance individuals' productive capacity, then, by definition, the job signalling model asserts that all individuals who complete education (presumably all individuals) are over-educated. This assumption, however, is unrealistic as years of economics research have provided much empirical evidence indicating the contrary, plus logic suggests that if education was indeed only a signal for allocating individuals to jobs then by now societies would surely have devised less expensive allocative mechanisms.

[9] Thurow (1975) argued that this approach was valid as empirical evidence for the US indicated the majority of skills used in the workplace came from on-the-job training (McGuinness, 2006). However, as previously mentioned, assuming education does not enhance individuals' labour productivity appears unrealistic. The job competition model, therefore, is unlikely to reflect the operations of modern labour markets.

[10] The job queue is formed as it is assumed individuals compete for high wage jobs, and the labour queue is formed because firms are assumed to compete for high productivity individuals (i.e., those with low training costs).

Given the assumption that education does not enhance individuals' labour productivity, then, by definition, the job competition model asserts that all individuals who complete education (presumably all individuals) are over-educated.[11] However, similar to in the job signalling model, completing this excess (unused) education is optimal for individuals because it directly affects their relative position in the labour queue, which ultimately determines the job (and wage) they obtain. The following scenario illustrates this point (Thurow, 1975; McGuinness, 2006). Suppose an individual observes a general rise in the educational attainment of other individuals in society; specifically, an increase in the proportion with a university degree. Based on human capital theory, this individual is less likely to complete such a degree as the increased supply of individuals with degrees would reduce the return to it. Based on the job competition model, however, this individual is more likely to complete a degree because it is a defensive tactic necessary to defend their position in the labour queue: the more individuals with a degree the more important it is to have one. Over-education, therefore, is an equilibrium outcome of the job competition model, and so it will be persistent for individuals (Linsley, 2005; McGuinness, 2006). Moreover, over-education will become more pronounced as more individuals complete higher education.

[11] That is, the skills (or knowledge) they acquired from education are not being used in their current job and they could have performed this job without the completion of their education.

Appendix 3.1: Details on the structure of ANZSCO

ANZSCO consists of five hierarchical levels: Major groups, Sub-major groups, Minor groups, Unit groups, and Occupations. Major groups are the broadest level of the classification and Occupations the most detailed. Table A3.1.1 provides an overview of the ANZSCO structure and coding.

Table A3.1.1: ANZSCO—Classification structure

Hierarchical level Details of level
Major groups • Broadest level of the classification • 1-digit codes used to denote • 8 Major groups in ANZSCO
Sub-major groups • Sub-divisions of the Major groups • 2-digit codes used to denote • 43 Sub-major groups in ANZSCO
Minor groups • Sub-divisions of the Sub-major groups • 3-digit codes used to denote • 97 Minor groups in ANZSCO
Unit groups • Sub-divisions of the Minor groups • 4-digit codes used to denote • 358 Unit groups in ANZSCO
Occupations • Sub-divisions of the Unit groups • 6-digit codes used to denote • 998 Occupations in ANZSCO

SOURCE: ABS (2006) ANZSCO, First Edition (Cat. no. 1220.0), page 11.

To further illustrate the structure of ANZSCO, Table A3.1.2 presents the titles and corresponding codes for each of the groups at the Major and Sub-major levels. Notice that among the groups at the Sub-major level there are 'not further defined (n.f.d)' categories. These 'n.f.d' categories (which are represented by codes ending with the digit '0') are supplementary codes used in the data coding process to categorise any inadequately described responses that cannot be coded to the specified level of the classification, but which can be coded to a higher (less detailed) level. Thus, 'n.f.d' categories, when they appear in the data, represent incomplete information within the occupation classification.

Table A3.1.2: ANZSCO—Major and Sub-major groups

Group code	Group title
1	**Managers**
10	Managers n.f.d
11	Chief Executives, General Managers and Legislators
12	Farmers and Farm Managers
13	Specialist Managers
14	Hospitality, Retail and Service Managers
2	**Professionals**
20	Professionals n.f.d
21	Arts and Media Professionals
22	Business, Human Resource and Marketing Professionals
23	Design, Engineering, Science and Transport Professionals
24	Education Professionals
25	Health Professionals
26	ICT Professionals
27	Legal, Social and Welfare Professionals
3	**Technicians and Trades Workers**
30	Technicians and Trades Workers n.f.d
31	Engineering, ICT and Science Technicians
32	Automotive and Engineering Trades Workers
33	Construction Trades Workers
34	Electrotechnology and Telecommunications Trades Workers
35	Food Trades Workers
36	Skilled Animal and Horticultural Workers
39	Other Technicians and Trades Workers
4	**Community and Personal Service Workers**
40	Community and Personal Service Workers n.f.d
41	Health and Welfare Support Workers
42	Carers and Aides
43	Hospitality Workers
44	Protective Service Workers
45	Sports and Personal Service Workers
5	**Clerical and Administrative Workers**
50	Clerical and Administrative Workers n.f.d
51	Office Managers and Program Administrators
52	Personal Assistants and Secretaries
53	General Clerical Workers
54	Inquiry Clerks and Receptionists
55	Numerical Clerks
56	Clerical and Office Support Workers
59	Other Clerical and Administrative Workers
6	**Sales Workers**
60	Sales Workers n.f.d
61	Sales Representatives and Agents
62	Sales Assistants and Salespersons
63	Sales Support Workers
7	**Machinery Operators and Drivers**
70	Machinery Operators and Drivers n.f.d
71	Machine and Stationary Plant Operators
72	Mobile Plant Operators
73	Road and Rail Drivers
74	Storepersons
8	**Labourers**
80	Labourers n.f.d
81	Cleaners and Laundry Workers
82	Construction and Mining Labourers
83	Factory Process Workers
84	Farm, Forestry and Garden Workers
85	Food Preparation Assistants
89	Other Labourers

SOURCE: ABS (2006) ANZSCO, First Edition (Cat. no. 1220.0), page 22.
NOTE: 'n.f.d' refers to 'not further defined' categories.

Appendix 3.2: Alterations to highest education level variables

Table A3.2.1: Cases where highest education level of individual decreases over time

Case (Occurrence)	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
1 (W1-W2)	Cert. I	Year 11	Year 11	Year 11	Year 11	Year 11	–	–
2 (W1-W2)	Year 12	Year 11	–	–	–	–	Year 11	Cert. IV
3 (W3-W4)	Year 10	Year 11	Year 12	Year 11	Year 11	–	Year 11	Year 11
4 (W5-W6)	Cert. II	Cert. II	Cert. II	Cert. II	Cert. II	Year 11	–	–
5 (W5-W6)	–	–	–	–	Cert. II	Year 11	Year 11	Year 11
6 (W6-W7)	Year 11	Year 12	Year 12	Year 12	Year 12	Year 12	Year 11	Year 12
7 (W6-W7)	Year 12	Year 12	–	Year 12	Year 12	Year 12	Year 11	Year 11

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTE: "–" indicates individual was a non-respondent in that particular wave of the HILDA Survey.

Since these cases likely represent measurement error, alterations are made to the highest education level variables; Table A3.2.2 outlines the alterations, whereby the original highest education level appears with ~~strikethrough~~ and the new highest education level appears below it.

Table A3.2.2: Alterations to highest education level variables

Case (Occurrence)	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
1 (W1-W2)	~~Cert. I~~ Year 11	Year 11	Year 11	Year 11	Year 11	Year 11	–	–
2 (W1-W2)	~~Year 12~~ Year 11	Year 11	–	–	–	–	Year 11	Cert. IV
3 (W3-W4)	Year 10	Year 11	~~Year 12~~ Year 11	Year 11	Year 11	–	Year 11	Year 11
4 (W5-W6)	Cert. II	Cert. II	Cert. II	Cert. II	Cert. II	~~Year 11~~ Cert. II	–	–
5 (W5-W6)	–	–	–	–	~~Cert. II~~ Year 11	Year 11	Year 11	Year 11
6 (W6-W7)	Year 11	Year 12	Year 12	Year 12	Year 12	Year 12	~~Year 11~~ Year 12	Year 12
7 (W6-W7)	Year 12	Year 12	–	Year 12	Year 12	Year 12	~~Year 11~~ Year 12	~~Year 11~~ Year 12

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTES: "–" indicates individual was a non-respondent in that particular wave of the HILDA Survey.

Appendix 3.3: Benchmarking measures of highest education level

Table A3.3.1: Benchmarking highest education levels, HILDA vs. ABS Survey of Education and Training (SET)—All individuals aged 15–64 years [a] (%)

ASCED levels of education	2001		2005	
	HILDA	SET	HILDA	SET
Postgraduate Degree or Graduate Diploma or Certificate [b]	6.8	6.0	7.4	6.8
Bachelor Degree	12.6	11.1	12.1	11.6
Advanced Diploma or Diploma [c]	8.4	7.6	8.2	8.4
Certificate III or IV	17.3	15.5	18.2	16.1
Certificate I or II	1.3	1.4	1.6	1.6
Certificate n.f.d [d]	0.5	0.2	0.6	0.1
Year 12	16.7	17.2	15.8	16.2
Year 11	7.5	8.3	5.7	6.9
Year 10 or below	28.9	31.4	30.4	31.2
Not determined [e]	0.0	1.3	0.0	1.1
Total	100.0	100.0	100.0	100.0

SOURCE: Author's calculations using HILDA Survey data (Release 8.0) and ABS Survey of Education and Training (SET) data (Cat. no. 6278.0).

NOTES: Figures are proportions that sum to 100.0 for each column; 2001 figures are representative of the Australian population of individuals who are between 15 and 64 years of age, while 2005 figures are representative of the Australian population of individuals who are 15 years of age or older: HILDA figures are weighted using cross-sectional population weights; reported SET figures already weighted.

(a) Reference population changes to individuals aged 15 years and over in 2005 due to a change in the scope of SET in 2005.

(b) Full title of category is 'Graduate Diploma or Graduate Certificate'.

(c) Category also includes Associate Degree (which is equivalent to an Advanced Diploma in ASCED).

(d) 'Certificate n.f.d' refers to 'Certificate not further defined'.

(e) SET figures for 'Not determined' include individuals who never attended school.

Table A3.3.2: Benchmarking highest education levels, HILDA vs. ABS Survey of Education and Work (SEW)—All individuals aged 15–64 years (%)

ASCED levels of education	2001		2002		2003		2004	
	HILDA	SEW	HILDA	SEW	HILDA	SEW	HILDA	SEW
Postgraduate Degree	2.6	2.2	2.7	2.3	3.1	2.5	3.2	2.8
Graduate Diploma or Certificate [a]	4.2	2.3	4.3	2.4	4.3	2.5	4.4	2.5
Bachelor Degree	12.6	12.5	12.5	13.1	12.5	13.0	12.8	13.6
Advanced Diploma or Diploma [b]	8.4	6.7	8.3	7.3	8.5	7.4	8.2	7.8
Certificate III or IV	17.3	14.1	17.7	14.8	17.9	15.2	18.4	15.5
Certificate I or II	1.3	1.0	1.4	0.9	1.5	1.0	1.6	0.9
Certificate n.f.d [c]	0.5	0.1	0.4	0.1	0.5	0.1	0.6	0.2
Year 12	16.7	19.7	16.6	19.8	17.3	20.1	17.6	20.0
Year 11	7.5	8.6	6.6	8.1	5.9	8.1	5.9	7.8
Year 10 or below	28.9	31.4	29.4	30.2	28.5	29.0	27.5	27.8
Not determined	0.0	1.1	0.2	0.6	0.1	0.9	0.0	0.7
Total [d]	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	2005		2006		2007		2008	
	HILDA	SEW	HILDA	SEW	HILDA	SEW	HILDA	SEW
Postgraduate Degree	3.3	2.9	3.4	3.2	3.6	3.4	3.9	3.9
Graduate Diploma or Certificate [a]	4.8	2.4	4.9	2.4	5.1	2.4	5.1	2.3
Bachelor Degree	13.2	14.2	13.6	15.0	13.6	14.9	14.0	15.8
Advanced Diploma or Diploma [b]	8.3	7.9	8.6	8.0	8.5	8.4	8.6	8.9
Certificate III or IV	18.8	14.7	18.9	15.4	19.4	15.9	19.5	16.1
Certificate I or II	1.6	1.0	1.7	0.9	1.7	0.9	1.8	0.9
Certificate n.f.d [c]	0.5	0.1	0.4	0.1	0.3	0.1	0.3	0.2
Year 12	17.4	20.4	17.7	20.3	17.5	20.5	17.6	19.7
Year 11	6.0	7.9	5.4	7.5	5.4	7.3	5.1	7.2
Year 10 or below	26.0	27.0	25.4	26.0	24.7	25.1	24.0	23.7
Not determined	0.0	1.2	0.0	1.1	0.1	0.7	0.0	1.2
Total [d]	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

SOURCE: Author's calculations using HILDA Survey data (Release 8.0) and ABS Survey of Education and Work (SEW) data (Cat. no. 6227.0).

NOTES: Figures are proportions that sum to 100.0 for (sub-sections in) each column and are representative of the Australian population of individuals who are between 15 and 64 years of age: HILDA figures are weighted using cross-sectional population weights; reported SEW figures already weighted.

(a) Full title of category is 'Graduate Diploma or Graduate Certificate'.

(b) Category also includes Associate Degree (which is equivalent to an Advanced Diploma in ASCED).

(c) 'Certificate n.f.d' refers to 'Certificate not further defined'.

(d) SEW figures for 'Total' include boarding school pupils and individuals who never attended school (which range from 0.1% to 0.4% across the years).

Appendix 3.4: Benchmarking ANZSCO occupation measures

Table A3.4.1: Benchmarking ANZSCO occupations at 1-digit level, HILDA vs. ABS Labour Force Survey (LFS)—All employed individuals aged 15 years and over (%)

ANZSCO 1-digit level occupation categories	2001		2002		2003		2004	
	HILDA	LFS	HILDA	LFS	HILDA	LFS	HILDA	LFS
Managers	12.8	12.3	13.0	12.2	12.4	12.0	12.7	12.4
Professionals	22.1	19.3	21.5	19.5	21.5	19.5	21.8	19.9
Technicians and Trades Workers	14.3	14.9	14.6	14.8	14.4	15.0	14.2	15.1
Community and Personal Service Workers	8.8	8.2	9.1	8.3	9.9	8.4	9.3	8.6
Clerical and Administrative Workers	15.1	16.6	14.9	16.2	14.9	16.2	15.3	15.7
Sales Workers	9.1	10.2	9.5	10.5	9.5	10.5	9.8	10.2
Machinery Operators and Drivers	6.7	6.9	6.6	6.8	6.3	6.7	6.3	6.7
Labourers	11.1	11.6	10.7	11.7	11.1	11.6	10.8	11.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	2005		2006		2007		2008	
	HILDA	LFS	HILDA	LFS	HILDA	LFS	HILDA	LFS
Managers	12.5	12.7	12.6	12.8	12.3	12.9	12.5	12.8
Professionals	22.7	19.8	22.8	20.3	22.6	20.3	22.7	20.9
Technicians and Trades Workers	14.3	15.2	14.5	15.2	14.7	15.2	14.2	15.3
Community and Personal Service Workers	9.1	8.6	9.6	8.5	10.1	8.7	10.2	8.7
Clerical and Administrative Workers	15.7	15.7	14.8	15.7	14.9	15.5	15.0	15.3
Sales Workers	9.8	10.5	9.2	10.0	9.4	9.5	9.0	9.5
Machinery Operators and Drivers	5.9	6.5	6.1	6.6	6.1	6.9	6.0	6.7
Labourers	10.0	11.2	10.4	11.1	9.9	11.0	10.5	10.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

SOURCE: Author's calculations using HILDA Survey data (Release 8.0) and ABS Labour Force Survey (LFS) data (Cat. no. 6291.0.55.003).

NOTES: Figures are proportions that sum to 100.0 for (sub-sections in) each column and are representative of the Australian population of employed individuals who are 15 years of age or older: HILDA figures are weighted using cross-sectional population weights; reported LFS figures already weighted.
Figures for ABS LFS are annual averages based on the published quarterly data.

Table A3.4.2: Benchmarking ANZSCO occupations at 2-digit level, HILDA vs. ABS Census of Population and Housing—All employed individuals aged 15 years and over (%)

ANZSCO 2-digit level occupation categories	2006	
	HILDA	Census
Managers n.f.d	0.0	0.5
Chief Executives, General Managers and Legislators	0.9	0.9
Farmers and Farm Managers	1.8	1.9
Specialist Managers	6.2	5.7
Hospitality, Retail and Service Managers	3.8	4.1
Professionals n.f.d	0.1	0.3
Arts and Media Professionals	1.2	0.8
Business, Human Resource and Marketing Professionals	5.5	4.9
Design, Engineering, Science and Transport Professionals	3.3	2.7
Education Professionals	5.6	4.4
Health Professionals	3.6	3.8
ICT Professionals	1.6	1.6
Legal, Social and Welfare Professionals	2.0	1.4
Technicians and Trades Workers n.f.d	0.0	0.2
Engineering, ICT and Science Technicians	2.1	1.9
Automotive and Engineering Trades Workers	3.3	3.3
Construction Trades Workers	2.7	3.0
Electrotechnology and Telecommunications Trades Workers	1.9	1.8
Food Trades Workers	1.3	1.4
Skilled Animal and Horticultural Workers	1.1	1.0
Other Technicians and Trades Workers	2.2	1.9
Community and Personal Service Workers n.f.d	0.0	0.0
Health and Welfare Support Workers	1.2	1.0
Carers and Aides	3.7	3.3
Hospitality Workers	2.1	2.0
Protective Service Workers	1.5	1.3
Sports and Personal Service Workers	1.0	1.2
Clerical and Administrative Workers n.f.d	0.0	0.0
Office Managers and Program Administrators	1.8	2.1
Personal Assistants and Secretaries	1.4	1.5
General Clerical Workers	2.6	2.9
Inquiry Clerks and Receptionists	2.5	2.2
Numerical Clerks	3.2	3.2
Clerical and Office Support Workers	0.9	1.0
Other Clerical and Administrative Workers	2.3	2.0
Sales Workers n.f.d	0.0	0.0
Sales Representatives and Agents	1.5	1.9
Sales Assistants and Salespersons	5.9	6.4
Sales Support Workers	1.9	1.5
Machinery Operators and Drivers n.f.d	0.0	0.1
Machine and Stationary Plant Operators	1.8	1.8
Mobile Plant Operators	1.0	1.0
Road and Rail Drivers	2.5	2.6
Storepersons	0.8	1.1
Labourers n.f.d	0.0	0.2
Cleaners and Laundry Workers	2.1	2.4
Construction and Mining Labourers	1.5	1.3
Factory Process Workers	2.2	2.3
Farm, Forestry and Garden Workers	0.8	1.1
Food Preparation Assistants	1.5	1.3
Other Labourers	2.3	1.9

Table A3.4.2 (continued)

ANZSCO 2-digit level occupation categories	2006	
	HILDA	Census
Not Determined [a]	0.0	1.8
Total	100.0	100.0

SOURCE: Author's calculations using HILDA Survey data (Release 8.0) and ABS Census of Population and Housing data (Cat. no. 2068.0).

NOTES: Figures are proportions that sum to 100.0 for each column and are representative of the Australian population of employed individuals who are 15 years of age or older: HILDA figures are weighted using cross-sectional population weights; reported Census figures are population estimates.
'n.f.d' refers to 'not further defined'.
(a) 'Not determined' figures consist of 'inadequately described' and 'not stated' cases.

Appendix 3.5: Further evidence from sample attrition analyses

Table A3.5.1: Probit model estimates for probability of non-response in any subsequent wave given individual responded in W1—Attrition from balanced panel W1–W8

Dependent variable: Non-response W2–W8	(I)		(II)		(III)	
	Coeff.	*SE*	*Coeff.*	*SE*	*Coeff.*	*SE*
Educational mismatch in W1 (base: Not employed)						
Over-educated	-0.115**	(0.037)	0.031	(0.042)	-0.410	(0.767)
Under-educated	0.017	(0.048)	0.104*	(0.052)	-0.392	(0.767)
Well-matched	-0.161**	(0.023)	0.074**	(0.029)	-0.394	(0.766)
Controls for basic individual characteristics	No		Yes		Yes	
Controls for extensive individual, household and interview characteristics	No		No		Yes	
N	13,969		13,969		13,969	
(No. non-respondents W2–W8)	(5,935)		(5,935)		(5,935)	
Pseudo R^2	0.0028		0.0572		0.0900	
Log likelihood	-9,497.56		-8,979.75		-8,666.90	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

See Notes to Table 3.8 for the list of variables included in the controls for basic individual characteristics and the controls for extensive individual, household and interview characteristics.

'Pseudo R^2' values equal 1 minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept term (Fitzgerald et al., 1998).

Complete sets of results are not presented, but are available from the author on request.

Table A3.5.2: Probit model estimates for probability of non-response at *t+1* given individual responded at *t* and *t-1*—Attrition from three-year panels across W1–W8

Dependent variable: Non-response $_{t+1}$	Models without additional control variables		Models with controls for individual, household and interview characteristics	
	Coeff.	*Robust SE*	*Coeff.*	*Robust SE*
A. Specification (I)				
Change in ANZSCO occupation category (base: No change)				
Changed occupation	0.114**	(0.025)	0.002	(0.027)
Not employed in both periods	0.054**	(0.015)	-0.024	(0.035)
Change in highest education level (base: No change)				
Increased highest education	0.122**	(0.035)	-0.060	(0.040)
Pseudo R^2	0.0071		0.0782	
Log likelihood	-18,591.29		-17,260.27	
B. Specification (II)				
Change in educational mismatch state (base: No change)				
Changed educational mismatch state	0.106**	(0.026)	-0.019	(0.028)
Not employed in both periods	0.050**	(0.015)	-0.032	(0.035)
Pseudo R^2	0.0066		0.0782	
Log likelihood	-18,601.20		-17,261.23	
C. Specification (III)				
Changes in educational mismatch state (base: No change)				
Change: Over-educated $_{t-1}$ – Under-educated $_t$	0.248*	(0.098)	0.049	(0.102)
Change: Over-educated $_{t-1}$ – Well-matched $_t$	0.111	(0.065)	-0.002	(0.069)
Change: Under-educated $_{t-1}$ – Over-educated $_t$	0.167	(0.124)	-0.016	(0.130)
Change: Under-educated $_{t-1}$ – Well-matched $_t$	0.105*	(0.043)	-0.015	(0.046)
Change: Well-matched $_{t-1}$ – Over-educated $_t$	0.103	(0.060)	0.001	(0.064)
Change: Well-matched $_{t-1}$ – Under-educated $_t$	0.069	(0.049)	-0.062	(0.051)
Not employed in both periods	0.050**	(0.015)	-0.031	(0.035)
Pseudo R^2	0.0067		0.0782	
Log likelihood	-18,599.73		-17,260.54	
N	69,618			
(No. non-respondents at *t+1*)	(5,293)			

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Controls for individual, household and interview characteristics are those included in specification (III) of the attrition models in Table 3.8 and Table A3.5.1; see Notes to Table 3.8 for full list of these controls.

All models also contain year dummies (i.e., identifiers for year which corresponds to period *t*).

'Pseudo R^2' values equal 1 minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept term (Fitzgerald et al., 1998).

Complete sets of results are not presented, but are available from the author on request.

Appendix 3.6: ANZSCO required education ranges of jobs

Table A3.6.1: ANZSCO required education ranges of jobs (and years of relevant experience to substitute for formal education)

ANZSCO skill level	Required education range (AQF qualifications levels)		Years of relevant experience to substitute for formal education
	Minimum	*Maximum*	
1	Bachelor Degree	Postgraduate Degree	5 years
2	Diploma	Advanced Diploma [a]	3 years
2–3	Certificate IV	Advanced Diploma [a]	3 years
3	Certificate IV	Certificate IV	3 years
3–4	Certificate II	Certificate IV	1 year
4	Certificate II	Certificate III	1 year
3–5	Year 10	Certificate IV	None
4–5	Year 10	Certificate III	None
5	Year 10	Certificate I	None

SOURCE: ABS (2006) ANZSCO, First Edition (Cat. no. 1220.0), pages 26-34.

NOTES: For skill level 3, since the number of years of on-the-job training each individual has completed in their current job cannot be reliably measured in the HILDA Survey data, it is assumed a Certificate IV is the sole required education level.

For skill level 5, it is assumed 'Compulsory Secondary School' refers to the completion of Year 10.

[a] Category also includes Associate Degree (which is equivalent to an Advanced Diploma in ASCED).

Appendix 4.1: Human capital and job requirements in Australia

Table A4.1.1: Highest education level by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

ASCED levels of education	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Postgraduate Degree	3.5	3.5	4.0	4.2	4.0	4.2	4.3	4.4
Graduate Diploma or Certificate [a]	6.2	5.9	5.6	6.0	6.6	6.3	6.4	6.5
Bachelor Degree	16.0	16.4	16.5	16.7	17.1	16.9	17.3	17.6
Advanced Diploma [b]	3.3	3.3	3.2	3.1	3.1	3.2	3.2	3.1
Diploma	6.4	6.1	6.3	6.1	6.4	6.7	6.8	6.9
Certificate IV	2.1	2.6	3.3	3.6	4.1	4.1	4.4	5.0
Certificate III	17.5	18.1	17.3	17.6	17.5	17.7	17.7	17.1
Certificate II	1.0	1.1	1.1	1.3	1.4	1.4	1.4	1.4
Certificate I	0.5	0.6	0.4	0.3	0.3	0.3	0.4	0.3
Year 12	17.1	16.4	16.7	16.8	16.6	16.9	16.5	17.1
Year 11	6.7	7.0	6.3	6.3	6.0	5.7	5.5	5.0
Year 10	13.8	13.6	13.1	12.7	11.9	11.6	11.5	10.8
Year 9 or below	6.0	5.4	6.1	5.4	5.0	5.0	4.7	4.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279	6,355

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

(a) Full title of category is 'Graduate Diploma or Graduate Certificate'.

(b) Category also includes Associate Degree (which is equivalent to an Advanced Diploma in ASCED).

Table A4.1.2: Labour market experience by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Total labour market experience (Time in paid work)								
None	0.5	0.7	0.5	0.4	0.5	0.4	0.4	0.3
Less than 1 year	1.7	2.3	2.6	3.4	2.1	2.1	2.3	2.2
1–2 years	5.2	5.4	6.8	6.9	6.5	6.3	5.5	5.1
3–5 years	8.6	8.7	8.6	8.0	9.1	9.7	9.5	10.2
6–9 years	11.7	12.0	11.4	11.8	11.0	11.2	11.2	11.6
10–19 years	28.9	28.2	27.0	26.6	27.7	26.8	27.1	26.8
20–29 years	24.7	25.0	24.8	24.7	24.0	24.2	23.8	23.1
30 or more years	18.7	17.7	18.4	18.2	19.0	19.3	20.2	20.7
Tenure in current occupation (i.e., relevant experience)								
Less than 1 year	14.5	16.7	16.5	17.9	18.1	16.8	18.1	17.6
1–2 years	17.3	15.9	16.7	16.7	16.4	16.7	15.3	16.7
3–5 years	18.5	18.9	20.2	19.0	20.1	18.5	19.7	18.8
6–9 years	14.0	14.3	13.1	13.9	13.2	15.1	13.2	13.9
10 or more years	35.7	34.3	33.6	32.6	32.2	33.0	33.7	33.0
Tenure with current employer								
Less than 1 year	22.0	22.3	21.4	22.6	22.3	21.4	22.7	21.4
1–2 years	22.3	22.5	21.4	20.2	20.4	21.8	19.2	20.0
3–5 years	19.4	19.7	21.8	22.9	22.2	20.6	21.5	20.4
6–9 years	12.4	12.9	13.0	12.3	12.7	14.1	13.0	14.3
10 or more years	23.9	22.5	22.4	22.1	22.5	22.2	23.6	23.9
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279	6,355

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for (sub-sections in) each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

Information on labour market experience is missing for some individuals; thus, reported proportions do not always correspond to the full sample. The extent of this problem is relatively minor for tenure in current occupation and tenure with current employer (less than 10 observations lost each wave), but not for total labour market experience (where number of observations lost steadily increases from roughly 80 in W3 to 400 in W8) due to increasing numbers of individuals with 'Not able to be determined' reported in HILDA derived variables for total labour market experience (`_ehtjb').

Table A4.1.3: Required education level of job (and years of relevant experience to substitute for formal education) by year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

ASCED levels of education	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)	Years of relevant experience to substitute for formal education
Bachelor Degree or higher	30.4	29.7	29.5	30.5	31.0	31.0	31.3	31.9	5 years
Diploma–Advanced Diploma	9.8	10.6	9.8	9.9	10.3	10.3	10.6	11.0	3 years
Certificate IV–Advanced Diploma	0.8	0.9	1.2	1.4	1.3	1.1	1.2	1.1	3 years
Certificate IV	13.4	13.8	13.6	12.5	12.5	12.9	12.9	11.9	3 years
Certificate II–Certificate IV	0.5	0.4	0.4	0.5	0.5	0.4	0.7	0.5	1 year
Certificate II–Certificate III	27.5	27.1	27.3	28.0	28.2	27.3	27.4	27.5	1 year
Year 10–Certificate IV	0.7	0.6	0.6	0.6	0.7	0.7	0.8	0.7	None
Year 10–Certificate III	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	None
Year 10–Certificate I	16.9	16.9	17.5	16.5	15.4	16.1	15.0	15.4	None
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279	6,355	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

Appendix 4.2: Over-education by population sub-groups

Table A4.2.1: Incidence of over-education by gender and year—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)	Overall (W1-W8)
Males									
Over-educated	16.2	18.3	19.3	18.5	18.2	19.4	18.7	19.6	18.6
Under-educated	10.3	11.4	11.4	12.5	11.9	12.5	12.1	12.2	11.8
Well-matched	73.5	70.3	69.4	69.0	69.9	68.1	69.2	68.2	69.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	3,141	3,006	3,034	2,952	3,060	3,127	3,112	3,165	24,597
Females									
Over-educated	21.5	21.3	22.5	22.4	23.4	24.1	24.1	23.8	23.0
Under-educated	8.1	9.6	8.5	8.9	9.4	9.2	9.5	8.3	9.0
Well-matched	70.4	69.1	68.9	68.7	67.2	66.8	66.4	67.9	68.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	3,077	2,864	2,904	2,865	3,060	3,131	3,167	3,190	24,258

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for (sub-sections in) each column and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

Table A4.2.2: Incidence of over-education by demographic characteristics and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

Demographics	Males				Females			
	Over.	Under.	W-M	*N*	Over.	Under.	W-M	*N*
Age								
15–19 years	27.4	38.5	34.2	*1,279*	34.3	18.5	47.3	*1,077*
20–24 years	26.4	19.1	54.5	*2,806*	33.0	13.7	53.3	*2,592*
25–34 years	21.1	11.9	67.0	*6,209*	25.2	10.0	64.8	*5,593*
35–44 years	15.0	10.1	74.9	*6,577*	21.7	8.6	69.7	*6,585*
45–54 years	15.3	7.2	77.5	*5,312*	18.1	6.3	75.6	*6,054*
55–64 years	15.0	5.0	80.0	*2,414*	16.5	4.6	78.9	*2,357*
Country of birth /Ethnicity								
Australia	17.2	12.9	69.9	*19,225*	20.6	9.4	70.0	*18,975*
ATSI	23.0	12.9	64.1	*419*	22.1	16.4	61.4	*427*
ESB migrant	12.9	10.7	76.5	*2,478*	21.6	8.9	69.5	*2,248*
NESB migrant	28.9	7.0	64.1	*2,473*	35.5	5.9	58.5	*2,608*
Marital status								
Married or de facto	15.4	9.7	74.9	*16,597*	21.0	8.1	70.9	*15,973*
Separated, divorced or widowed	20.8	7.4	71.8	*1,647*	19.6	8.9	71.5	*3,251*
Single	24.7	17.1	58.3	*6,350*	29.9	11.2	58.9	*5,028*
Long term health condition, disability or impairment								
Yes	19.5	9.8	70.7	*3,177*	24.7	9.3	65.9	*3,169*
No	18.4	12.1	69.5	*21,394*	22.7	8.9	68.4	*21,072*
Remoteness of area of residence								
Major city	18.8	11.6	69.6	*15,915*	24.0	8.9	67.1	*15,856*
Inner regional	16.8	12.1	71.1	*5,682*	18.5	8.7	72.8	*5,476*
Outer regional	21.7	12.5	65.8	*2,503*	24.2	9.6	66.2	*2,432*
Remote	11.9	13.4	74.7	*388*	22.7	9.4	67.9	*392*
Very remote	13.0	17.6	69.4	*108*	18.6	6.9	74.5	*102*
State/Territory of residence								
NSW	18.9	11.0	70.0	*7,172*	21.8	7.8	70.4	*7,211*
Victoria	19.0	11.1	69.9	*6,089*	25.4	9.3	65.3	*6,079*
Queensland	17.3	13.9	68.8	*5,351*	20.7	10.0	69.3	*5,142*
South Australia	20.3	10.6	69.0	*2,181*	23.6	10.4	66.0	*2,161*
Western Australia	18.4	12.7	68.9	*2,338*	25.3	8.7	66.0	*2,088*
Tasmania	16.2	9.9	73.9	*671*	20.6	4.9	74.5	*800*
NT	26.7	15.6	57.7	*211*	35.2	9.0	55.8	*231*
ACT	13.3	12.5	74.2	*584*	17.3	13.8	68.9	*546*
Overall	18.6	11.8	69.6	*24,597*	23.0	9.0	68.1	*24,258*
N	24,597				24,258			

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each row (separately for males and females) (except 'N' columns, which report number of observations in each particular row) and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed for 2001–2008 period.

Information missing for some individuals; thus, reported proportions do not always correspond to the full sample. The extent of this problem is relatively minor (with less than 30 observations lost in each instance).

'ATSI' refers to Aboriginal or Torres Strait Islander background; 'ESB' refers to English speaking background; 'NESB' refers to non-English speaking background.

Table A4.2.3: Incidence of over-education by human capital measures and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

Human capital	Males				Females			
	Over.	Under.	W-M	*N*	Over.	Under.	W-M	*N*
Highest education level								
Postgraduate Degree	14.4	0.0	85.6	*1,141*	17.9	0.0	82.1	*871*
Graduate Diploma or Certificate	19.3	0.0	80.7	*1,289*	17.8	0.0	82.2	*2,022*
Bachelor Degree	29.9	0.0	70.1	*3,549*	30.7	0.0	69.3	*4,561*
Advanced Diploma	34.1	16.4	49.5	*583*	28.7	10.3	61.1	*940*
Diploma	41.3	11.6	47.1	*1,535*	57.7	11.9	30.4	*1,579*
Certificate IV	33.8	16.5	49.7	*920*	54.6	15.0	30.4	*916*
Certificate III	11.1	13.2	75.7	*6,001*	16.7	12.4	70.8	*2,816*
Certificate II	26.2	10.3	63.4	*271*	26.5	8.8	64.8	*306*
Certificate I	0.0	13.1	86.9	*87*	0.0	7.7	92.3	*89*
Year 12	20.8	21.3	57.9	*3,778*	23.7	14.6	61.7	*4,077*
Year 11	26.8	13.2	60.0	*1,479*	28.0	7.3	64.6	*1,568*
Year 10	0.0	16.1	83.9	*2,577*	0.0	14.8	85.2	*3,355*
Year 9 or below	0.0	16.1	83.9	*1,387*	0.0	11.3	88.7	*1,158*
Total labour market experience (Time in paid work)								
None	49.1	18.6	32.3	*99*	40.1	16.7	43.2	*110*
Less than 1 year	39.0	30.9	30.1	*565*	44.6	15.5	39.9	*520*
1–2 years	30.5	31.3	38.3	*1,372*	34.9	14.0	51.1	*1,382*
3–5 years	25.6	16.4	58.0	*1,994*	26.8	10.6	62.7	*2,133*
6–9 years	22.6	10.9	66.5	*2,464*	26.5	9.4	64.2	*2,692*
10–19 years	17.2	9.6	73.2	*5,884*	22.7	9.6	67.7	*7,265*
20–29 years	14.1	10.1	75.8	*6,064*	17.6	6.6	75.8	*6,045*
30 or more years	13.8	6.4	79.9	*5,294*	16.3	5.1	78.7	*3,318*
Tenure in current occupation (i.e., relevant experience)								
Less than 1 year	28.0	34.3	37.7	*4,033*	32.1	24.5	43.4	*4,544*
1–2 years	25.9	30.6	43.5	*3,872*	27.9	19.4	52.6	*4,163*
3–5 years	21.5	8.1	70.4	*4,618*	25.5	5.6	68.8	*4,649*
6–9 years	15.4	0.0	84.6	*3,324*	21.3	0.0	78.7	*3,162*
10 or more years	10.7	0.0	89.3	*8,731*	13.6	0.0	86.4	*7,720*
Tenure with current employer								
Less than 1 year	24.6	22.3	53.0	*5,375*	30.4	16.2	53.3	*5,678*
1–2 years	21.7	19.2	59.1	*5,071*	25.8	13.1	61.1	*5,176*
3–5 years	19.4	7.8	72.7	*5,034*	23.1	6.5	70.4	*5,129*
6–9 years	15.8	3.6	80.6	*3,083*	20.3	2.9	76.8	*3,126*
10 or more years	11.2	3.9	84.9	*6,016*	13.4	3.4	83.2	*5,124*
Overall	18.6	11.8	69.6	*24,597*	23.0	9.0	68.1	*24,258*
N	24,597				24,258			

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each row (separately for males and females) (except 'N' columns, which report number of observations in each particular row) and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed for 2001–2008 period.

Information on labour market experience is missing for some individuals; thus, reported proportions do not always correspond to the full sample. The extent of this problem is relatively minor for tenure in current occupation and tenure with current employer (less than 20 observations lost in each instance), but not for total labour market experience (where roughly 800 observations are lost for males and females) due to individuals with 'Not able to be determined' reported in HILDA derived variables for total labour market experience ('_ehtjb').

Table A4.2.4: Incidence of over-education by job characteristics and gender—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

Job characteristics	Males				Females			
	Over.	Under.	W-M	*N*	Over.	Under.	W-M	*N*
Occupation (ANZSCO 1-digit level)								
Managers	4.6	22.6	72.8	*3,201*	6.2	25.9	67.9	*1,806*
Professionals	0.0	13.2	86.8	*5,114*	0.1	10.2	89.7	*6,858*
Technicians and Trades Workers	10.5	21.0	68.5	*5,394*	22.4	23.1	54.6	*992*
Community and Personal Service Workers	21.9	6.5	71.6	*1,580*	26.6	5.7	67.6	*3,608*
Clerical and Administrative Workers	41.2	5.3	53.5	*2,058*	26.8	7.9	65.3	*6,209*
Sales Workers	52.4	2.6	45.0	*1,405*	61.9	3.0	35.1	*2,572*
Machinery Operators and Drivers	7.7	4.5	87.8	*2,910*	4.8	4.3	90.9	*260*
Labourers	55.8	0.8	43.4	*2,935*	46.3	0.2	53.5	*1,953*
Industry (ANZSIC Division level)								
Agriculture, Forestry and Fishing	29.0	8.3	62.7	*804*	33.8	4.7	61.5	*247*
Mining	9.6	10.2	80.1	*765*	20.5	14.8	64.7	*97*
Manufacturing	17.1	12.3	70.6	*4,071*	23.5	8.4	68.1	*1,341*
Electricity, Gas, Water and Waste Services	9.4	12.9	77.7	*456*	11.6	11.3	77.1	*104*
Construction	13.7	20.3	66.1	*2,319*	25.1	12.2	62..6	*328*
Wholesale Trade	15.3	12.4	72.3	*1,159*	28.0	11.1	60.9	*600*
Retail Trade	37.9	8.9	53.1	*1,839*	51.2	5.7	43.2	*2,610*
Accommodation and Food Services	28.3	10.0	61.7	*1,021*	30.6	10.3	59.1	*1,580*
Transport, Postal and Warehousing	22.4	5.8	71.8	*1,680*	28.1	7.3	64.6	*536*
Information Media and Telecommunications	16.5	21.4	62.1	*705*	25.4	12.8	61.8	*620*
Financial and Insurance Services	16.9	12.3	70.8	*811*	23.8	12.8	63.4	*1,151*
Rental, Hiring and Real Estate Services	32.5	16.1	51.4	*258*	18.8	20.7	60.5	*361*
Professional, Scientific and Technical Services	9.6	12.3	78.1	*1,591*	16.6	14.4	69.1	*1,636*
Administrative and Support Services	27.6	19.0	53.4	*454*	28.5	14.9	56.6	*732*
Public Administration and Safety	16.9	9.7	73.4	*2,464*	25.0	11.3	63.6	*1,599*
Education and Training	7.4	6.6	86.0	*1,602*	9.9	4.3	85.7	*4,058*
Health Care and Social Assistance	17.9	6.3	75.8	*1,098*	14.3	6.8	78.9	*5,547*
Arts and Recreation Services	34.7	9.6	55.7	*474*	29.8	13.1	57.1	*336*
Other Services	13.5	18.3	68.2	*951*	22.4	15.8	61.7	*703*
Type of employer / Sector of employment								
Private, For Profit	19.4	12.8	67.8	*17,974*	27.7	10.2	62.1	*13,915*
Government Business Enterprise	18.1	11.2	70.8	*1,632*	22.2	9.7	68.1	*1,387*
Other Commercial Organisation	21.4	8.4	70.2	*75*	27.4	3.7	68.9	*84*
Private, Not-For-Profit	17.2	12.9	69.9	*936*	18.2	9.0	72.8	*2,153*
Other Government Organisation	15.2	6.8	77.9	*3,801*	13.5	5.7	80.8	*6,300*
Other Non-commercialOrganisation	13.8	12.6	73.6	*163*	17.7	9.6	72.7	*381*
Employer size (Number employed at place of work)								
Less than 5	19.0	18.2	62.8	*2,555*	24.8	11.5	63.7	*2,347*
5–9	20.4	14.0	65.6	*2,884*	28.9	12.0	59.0	*3,012*
10–19	19.4	12.5	68.1	*3,448*	24.7	9.5	65.8	*3,424*
20–49	18.5	12.5	69.0	*4,323*	19.0	9.1	71.9	*4,482*
50–99	18.3	10.8	70.9	*2,998*	21.4	7.2	71.4	*3,054*
100–199	18.4	8.8	72.8	*2,638*	24.3	6.3	69.4	*2,371*
200–499	18.6	10.0	71.4	*2,452*	24.4	8.2	67.4	*2,201*
500 or more	15.3	8.1	76.6	*3,091*	18.8	7.6	73.6	*2,983*
Unsure, but less than 20	38.5	15.1	46.4	*31*	27.4	13.1	59.5	*51*
Unsure, but 20 or more	27.7	13.7	58.6	*171*	23.9	8.2	67.9	*328*

Table A4.2.4 (continued)

Job characteristics	Males				Females			
	Over.	Under.	W-M	*N*	Over.	Under.	W-M	*N*
Usual hours of work								
Full-time (35 or more per week)	16.4	12.2	71.4	*22,055*	19.3	10.6	70.1	*13,729*
Part-time (Less than 35 per week)	37.0	8.6	54.3	*2,542*	28.1	6.7	65.2	*10,529*
Employment contract type								
Fixed term	13.4	17.5	69.1	*2,297*	19.5	11.8	68.7	*2,377*
Casual	35.9	12.1	51.9	*3,482*	34.7	8.2	57.0	*5,681*
Permanent or ongoing	16.1	11.1	72.9	*18,714*	19.5	8.8	71.7	*16,104*
Other	17.9	21.1	60.9	*91*	29.4	8.1	62.4	*84*
Union member								
Yes	17.2	6.6	76.3	*7,430*	16.0	4.5	79.5	*6,939*
No	19.2	14.1	66.7	*17,158*	25.7	10.7	63.6	*17,301*
Some of usual hours worked at home								
Yes	9.2	11.4	79.4	*4,486*	11.6	9.2	79.2	*4,646*
No	20.5	11.9	67.6	*20,110*	25.6	8.9	65.6	*19,611*
Overall	18.6	11.8	69.6	*24,597*	23.0	9.0	68.1	*24,258*
N	24,597				24,258			

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each row (separately for males and females) (except 'N' columns, which report number of observations in each particular row) and are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed for 2001–2008 period.

Information missing for some individuals; thus, reported proportions do not always correspond to the full sample. The extent of this problem is relatively minor (less than 80 observations lost in each instance).

Appendix 5.1: Definition of variables and descriptive statistics

Table A5.1.1: Definition of dependent and independent variables used in analyses

Variable name	Description
Real weekly wage	Individual's usual weekly gross wages from their current (main) job (based on HILDA derived variable '_wscmei'), CPI-adjusted to be expressed in 2001-dollars in each year (based on ABS Consumer Price Index information for eight capital cities of Australia (Cat. no. 6401.0))
Real hourly wage	Individual's usual hourly wage rate, CPI-adjusted to be expressed in 2001-dollars in each year (calculated using their real weekly wage and usual weekly hours worked in current (main) job (HILDA derived variable '_jbmhruc'))
ln(real hourly wage)	Natural logarithm of individual's real hourly wage
Postgraduate Degree	Highest education level: Postgraduate Degree (Masters or Doctorate)
Graduate Diploma or Certificate	Highest education level: Graduate Diploma or Graduate Certificate
Bachelor Degree	Highest education level: Bachelor Degree (including Bachelor (Honours) Degree)
Advanced Diploma	Highest education level: Advanced Diploma or Associate Degree
Diploma	Highest education level: Diploma
Certificate IV	Highest education level: Certificate IV
Certificate III	Highest education level: Certificate III
Certificate II	Highest education level: Certificate II
Certificate I	Highest education level: Certificate I
Year 12	Highest education level: Year 12 (Senior Secondary education)
Year 11	Highest education level: Year 11
Year 10 or below	Highest education level: Year 10 or below
Multiple qualifications	Individual has completed more than one post-school qualification
Experience	Years spent in paid employment
Occupation tenure	Years spent working in current occupation (i.e., relevant experience)
Employer tenure	Years spent working for current employer
Years unemployed	Years spent unemployed and looking for work
Years NILF	Years spent not-in-the-labour-force (NILF)
English only	English proficiency: English is only language spoken
Good	English proficiency: English spoken very well or well
Poor	English proficiency: English spoken not well
L-T health condition	Individual has a long-term health condition, disability or impairment (which has or will last for six months or more)
General health rating	Rating of individual's general health based on SF-36 Health Survey instrument in the Self-Completion Questionnaire (SCQ) for each wave of the HILDA Survey (HILDA derived variable '_ghgh') (measured on a 0–100 index)
Mental health rating	Rating of individual's mental health based on SF-36 Health Survey instrument in the SCQ for each wave of the HILDA Survey (HILDA derived variable '_ghmh') (measured on a 0–100 index)
% time employed last year	Proportion of time spent employed during last financial year
% time unemp. last year	Proportion of time spent unemployed during last financial year
% time NILF last year	Proportion of time spent NILF during last financial year
No. jobs last year	Number of jobs had during last financial year
15–19 years	Age: 15 to 19 years
20–24 years	Age: 20 to 24 years
25–29 years	Age: 25 to 29 years
30–34 years	Age: 30 to 34 years
35–39 years	Age: 35 to 39 years
40–44 years	Age: 40 to 44 years

Table A5.1.1 (continued)

Variable name	Description
45–49 years	Age: 45 to 49 years
50–54 years	Age: 50 to 54 years
55–59 years	Age: 55 to 59 years
60–64 years	Age: 60 to 64 years
Australian-born	Country of birth/Ethnicity: Australian-born, non-ATSI
ATSI	Country of birth/Ethnicity: Aboriginal or Torres Strait Islander (ATSI)
Migrant ESB	Country of birth/Ethnicity: Migrant from main English-speaking background (ESB) country (US, Canada, South Africa, UK, Ireland, and New Zealand)
Migrant Other	Country of birth/Ethnicity: Migrant from other (non-main ESB) (NESB) country
Legally married	Marital status: Legally married
Defacto	Marital status: Defacto
Separated	Marital status: Separated
Divorced	Marital status: Divorced
Widowed	Marital status: Widowed
Single	Marital status: Single
No. children had	Total number of children ever had
Resident children aged 0–4 years	Number of (own) resident children aged 0 to 4 years in household (HH)
Resident children aged 5–14 years	Number of (own) resident children aged 5 to 14 years in HH
Resident children aged 15–24 years	Number of (own) resident children aged 15 to 24 years in HH
Resident children aged 25+ years	Number of (own) resident children aged 25 years or older in HH
Non-resident children aged 0–4 years	Number of (own) non-resident children aged 0 to 4 years in HH
Non-resident children aged 5–14 years	Number of (own) non-resident children aged 5 to 14 years in HH
Non-resident children aged 15–24 years	Number of (own) non-resident children aged 15 to 24 years in HH
Non-resident children aged 25+ years	Number of (own) non-resident children aged 25 years or older in HH
Rest HH income	Total income earned by all other members of individual's HH (calculated as total HH financial year gross income (HILDA derived variables '_hifefp', '_hifefn') minus individual's financial year gross income (HILDA derived variables '_tifefp', '_tifefn')), CPI-adjusted to be expressed in 2001-dollars in each year (based on ABS Consumer Price Index information for eight capital cities (Cat. no. 6401.0))
NSW	State/Territory of residence: New South Wales (NSW)
Victoria	State/Territory of residence: Victoria
Queensland	State/Territory of residence: Queensland
South Australia	State/Territory of residence: South Australia
Western Australia	State/Territory of residence: Western Australia
Tasmania	State/Territory of residence: Tasmania
NT	State/Territory of residence: Northern Territory (NT)
ACT	State/Territory of residence: Australian Capital Territory (ACT)
Major city	Remoteness of area of residence: Major city of Australia
Inner regional	Remoteness of area of residence: Inner regional Australia
Outer regional	Remoteness of area of residence: Outer regional Australia
Remote (very remote)	Remoteness of area of residence: Remote or very remote Australia
Lived in 1 home	Number homes lived in during past ten years: 1 home
Lived in 2–3 homes	Number homes lived in during past ten years: 2 to 3 homes
Lived in 4–9 homes	Number homes lived in during past ten years: 4 to 9 homes
Lived in 10+ homes	Number homes lived in during past ten years: 10 or more homes
Father Australian-born	Father's country of birth/ethnicity: Australian-born (including ATSI)
Father Migrant ESB	Father's country of birth/ethnicity: Migrant from ESB country
Father Migrant Other	Father's country of birth/ethnicity: Migrant from other NESB country
Mother Australian-born	Mother's country of birth/ethnicity: Australian-born (including ATSI)
Mother Migrant ESB	Mother's country of birth/ethnicity: Migrant from ESB country
Mother Migrant Other	Mother's country of birth/ethnicity: Migrant from other NESB country
Father not employed when aged 14	Individual's father was not employed when individual was age 14

Table A5.1.1 (continued)

Variable name	Description
Father employed when aged 14	Individual's father was employed when individual was age 14
Father's employment status unknown	Individual's father's employment status when individual was age 14 unknown because either father was deceased, individual was not living with father or individual is unable to determine
Father unemployed for 6+ months	Individual's father was unemployed for six months or more while the individual was growing up
Father not unemployed for 6+ months	Individual's father was not unemployed for six months or more while the individual was growing up
Father's unemployment history unknown	Individual's father's unemployment history is unknown because either father was deceased or individual is unable to determine
Mother not employed when aged 14	Individual's mother was not employed when individual was age 14
Mother employed when aged 14	Individual's mother was employed when individual was age 14
Mother's employment status unknown	Individual's mother's employment status when individual was age 14 unknown because either mother was deceased, individual was not living with mother or individual is unable to determine
No. siblings	Total number of siblings
Eldest child	Individual is the eldest child among his/her siblings
Father had post-school qualification	Individual's father completed a post-school qualification (variable derived in wave 5 of the HILDA Survey data, then applied to all other waves)
Mother had post-school qualification	Individual's mother completed a post-school qualification (variable derived in wave 5 of the HILDA Survey data, then applied to all other waves)
Openness	Personality measure: Openness to experience – HILDA derived variable for individual's personality character traits ('_pnopene') (variable derived in wave 5 SCQ, then applied to all other waves) (scale 1–7, with higher values meaning trait better describes individual) (see Summerfield (2010) for further details)
Conscientiousness	Personality measure: Conscientiousness – HILDA derived variable for individual's personality character traits ('_pnconsc') (variable derived in wave 5 SCQ, then applied to all other waves) (scale –7, with higher values meaning trait better describes individual) (see Summerfield (2010) for further details)
Extroversion	Personality measure: Extroversion – HILDA derived variable for individual's personality character traits ('_pnextrv') (variable derived in wave 5 SCQ, then applied to all other waves) (scale 1–7, with higher values meaning trait better describes individual) (see Summerfield (2010) for further details)
Agreeableness	Personality measure: Agreeableness – HILDA derived variable for individual's personality character traits ('_pnagree') (variable derived in wave 5 SCQ, then applied to all other waves) (scale 1–7, with higher values meaning trait better describes individual) (see Summerfield (2010) for further details)
Emotional stability	Personality measure: Emotional stability – HILDA derived variable for individual's personality character traits ('_pnemote') (variable derived in wave 5 SCQ, then applied to all other waves) (scale 1–7, with higher values meaning trait better describes individual) (see Summerfield (2010) for further details)

SOURCE: Summerfield (2010) and the HILDA Online Data Dictionary (see http://www.melbourneinstitute.com/hildaddictionary/onlinedd/default.aspx).

Table A5.1.2: Descriptive statistics for dependent and independent variables by over-education status and gender

Variable	Males					Females				
	Over-educated		Well-matched		*N*	Over-educated		Well-matched		*N*
	mean	s.d	mean	s.d		mean	s.d	mean	s.d	
Real weekly wage	710.22	491.25	974.08	596.93	*21,617*	465.99	302.35	628.27	379.54	*22,042*
Real hourly wage	18.41	16.05	22.52	15.57	*21,587*	16.47	12.16	19.60	13.28	*22,014*
ln(Real hourly wage)	2.800	0.483	2.996	0.496	*21,459*	2.704	0.439	2.875	0.458	*21,850*
Highest education level										
Postgraduate Degree	0.036	0.186	0.057	0.232	*21,617*	0.023	0.150	0.045	0.207	*22,042*
Graduate Diploma or Certificate	0.055	0.227	0.061	0.239	*21,617*	0.066	0.248	0.100	0.300	*22,042*
Bachelor Degree	0.226	0.418	0.149	0.356	*21,617*	0.248	0.432	0.194	0.395	*22,042*
Advanced Diploma	0.046	0.210	0.017	0.130	*21,617*	0.048	0.213	0.035	0.185	*22,042*
Diploma	0.134	0.341	0.044	0.204	*21,617*	0.167	0.373	0.029	0.169	*22,042*
Certificate IV	0.068	0.252	0.027	0.163	*21,617*	0.090	0.286	0.017	0.131	*22,042*
Certificate III	0.150	0.357	0.263	0.440	*21,617*	0.083	0.276	0.121	0.326	*22,042*
Certificate II	0.014	0.117	0.010	0.099	*21,617*	0.018	0.133	0.011	0.104	*22,042*
Certificate I	0.000	0.000	0.004	0.065	*21,617*	0.000	0.000	0.005	0.069	*22,042*
Year 12	0.183	0.387	0.125	0.331	*21,617*	0.176	0.381	0.151	0.358	*22,042*
Year 11	0.088	0.283	0.053	0.223	*21,617*	0.081	0.273	0.060	0.237	*22,042*
Year 10 or below	0.000	0.000	0.190	0.392	*21,617*	0.000	0.000	0.231	0.422	*22,042*
Multiple qualifications	0.402	0.490	0.355	0.479	*21,617*	0.435	0.496	0.377	0.485	*22,042*
Experience	16.73	12.44	21.23	11.83	*20,866*	14.74	10.42	18.64	10.53	*21,306*
Occupation tenure	5.78	7.53	11.23	9.96	*21,606*	5.21	6.41	9.95	9.49	*22,025*
Employer tenure	4.77	6.75	7.75	8.42	*21,606*	3.88	4.79	6.75	7.40	*22,020*
Years unemployed	0.79	1.72	0.53	1.34	*20,845*	0.55	1.53	0.38	1.31	*21,273*
Years NILF	1.19	2.58	0.90	1.98	*20,840*	3.85	5.17	4.29	5.62	*21,272*
English proficiency										
English only	0.868	0.338	0.916	0.277	*21,616*	0.852	0.355	0.922	0.268	*22,042*
Good	0.120	0.325	0.079	0.269	*21,616*	0.137	0.344	0.075	0.263	*22,042*
Poor	0.011	0.106	0.005	0.071	*21,616*	0.010	0.102	0.003	0.057	*22,042*
L-T health condition	0.140	0.347	0.129	0.335	*21,598*	0.136	0.343	0.128	0.334	*22,028*
General health rating	71.99	18.00	71.91	18.05	*19,302*	71.62	18.94	73.41	18.68	*20,183*
Mental health rating	75.97	15.64	76.29	15.39	*19,397*	72.88	16.44	74.77	15.75	*20,344*
% time employed last year	91.41	22.20	96.94	13.20	*21,617*	89.40	24.80	94.61	18.22	*22,042*
% time unemp. last year	4.33	15.41	1.55	8.98	*21,617*	3.82	14.38	1.76	9.75	*22,042*
% time NILF last year	4.26	15.73	1.50	9.26	*21,617*	6.78	20.21	3.63	14.93	*22,042*
No. jobs last year	1.35	0.77	1.24	0.59	*21,617*	1.35	0.76	1.26	0.62	*22,042*
Age										
15–19 years	0.079	0.270	0.026	0.160	*21,617*	0.066	0.248	0.031	0.172	*22,042*
20–24 years	0.159	0.365	0.091	0.287	*21,617*	0.151	0.359	0.086	0.280	*22,042*
25–29 years	0.140	0.347	0.114	0.317	*21,617*	0.129	0.336	0.105	0.307	*22,042*
30–34 years	0.130	0.337	0.134	0.341	*21,617*	0.129	0.335	0.113	0.316	*22,042*
35–39 years	0.115	0.319	0.141	0.348	*21,617*	0.131	0.338	0.127	0.333	*22,042*
40–44 years	0.109	0.311	0.145	0.352	*21,617*	0.122	0.328	0.152	0.359	*22,042*
45–49 years	0.102	0.303	0.133	0.340	*21,617*	0.115	0.319	0.152	0.359	*22,042*
50–54 years	0.081	0.273	0.106	0.307	*21,617*	0.084	0.277	0.123	0.329	*22,042*
55–59 years	0.054	0.226	0.073	0.261	*21,617*	0.053	0.224	0.078	0.268	*22,042*
60–64 years	0.031	0.174	0.037	0.190	*21,617*	0.020	0.139	0.033	0.180	*22,042*
Country of birth/Ethnicity										
Australian-born	0.752	0.432	0.781	0.414	*21,617*	0.719	0.450	0.800	0.400	*22,042*
ATSI	0.017	0.131	0.016	0.125	*21,617*	0.018	0.133	0.015	0.121	*22,042*
Migrant ESB	0.077	0.267	0.108	0.311	*21,617*	0.090	0.286	0.094	0.292	*22,042*
Migrant Other	0.153	0.360	0.095	0.293	*21,617*	0.173	0.379	0.091	0.288	*22,042*

Table A5.1.2 (continued)

Variable	Males					Females				
	Over-educated		Well-matched		*N*	Over-educated		Well-matched		*N*
	mean	s.d	mean	s.d		mean	s.d	mean	s.d	
Marital status										
Legally married	0.413	0.492	0.563	0.496	*21,614*	0.433	0.495	0.531	0.499	*22,037*
Defacto	0.159	0.366	0.157	0.364	*21,614*	0.182	0.386	0.147	0.355	*22,037*
Separated	0.030	0.169	0.025	0.155	*21,614*	0.034	0.182	0.039	0.194	*22,037*
Divorced	0.048	0.214	0.040	0.195	*21,614*	0.068	0.251	0.083	0.276	*22,037*
Widowed	0.000	0.015	0.004	0.059	*21,614*	0.011	0.103	0.019	0.137	*22,037*
Single	0.351	0.477	0.212	0.409	*21,614*	0.273	0.445	0.180	0.385	*22,037*
No. children had	1.08	1.35	1.46	1.40	*21,615*	1.24	1.33	1.54	1.38	*22,042*
Resident children by age										
No. aged 0–4 years	0.18	0.49	0.22	0.53	*21,617*	0.17	0.46	0.14	0.43	*22,042*
No. aged 5–14 years	0.26	0.67	0.42	0.82	*21,617*	0.41	0.78	0.45	0.82	*22,042*
No. aged 15–24 years	0.16	0.50	0.23	0.59	*21,617*	0.22	0.56	0.32	0.67	*22,042*
No. aged 25+ years	0.02	0.14	0.02	0.16	*21,617*	0.03	0.19	0.04	0.21	*22,042*
Non-resident children by age										
No. aged 0–4 years	0.02	0.15	0.02	0.15	*21,609*	0.00	0.04	0.00	0.02	*22,037*
No. aged 5–14 years	0.09	0.40	0.09	0.39	*21,609*	0.01	0.13	0.01	0.11	*22,037*
No. aged 15–24 years	0.14	0.49	0.18	0.54	*21,609*	0.13	0.44	0.18	0.50	*22,037*
No. aged 25+ years	0.20	0.65	0.26	0.73	*21,609*	0.24	0.71	0.37	0.86	*22,037*
Rest HH income	34214.2	39225.9	29304.8	32343.2	*21,617*	47572.7	51150.7	46598.9	49391.4	*22,042*
State/Territory of residence										
NSW	0.288	0.453	0.296	0.456	*21,617*	0.280	0.449	0.306	0.461	*22,042*
Victoria	0.243	0.429	0.252	0.434	*21,617*	0.274	0.446	0.244	0.429	*22,042*
Queensland	0.219	0.414	0.211	0.408	*21,617*	0.206	0.405	0.211	0.408	*22,042*
South Australia	0.094	0.292	0.089	0.284	*21,617*	0.090	0.287	0.089	0.284	*22,042*
Western Australia	0.099	0.298	0.093	0.291	*21,617*	0.088	0.284	0.085	0.279	*22,042*
Tasmania	0.025	0.157	0.029	0.167	*21,617*	0.031	0.174	0.035	0.184	*22,042*
NT	0.011	0.103	0.007	0.085	*21,617*	0.013	0.113	0.008	0.092	*22,042*
ACT	0.020	0.140	0.024	0.153	*21,617*	0.018	0.131	0.022	0.147	*22,042*
Remoteness of area of residence										
Major city	0.652	0.476	0.648	0.478	*21,616*	0.677	0.468	0.647	0.478	*22,042*
Inner regional	0.217	0.412	0.235	0.424	*21,616*	0.197	0.398	0.235	0.424	*22,042*
Outer regional	0.114	0.318	0.097	0.296	*21,616*	0.106	0.308	0.098	0.297	*22,042*
Remote (very remote)	0.017	0.129	0.020	0.141	*21,616*	0.019	0.137	0.020	0.142	*22,042*
No. homes lived in										
Lived in 1 home	0.362	0.481	0.328	0.469	*21,603*	0.309	0.462	0.353	0.478	*22,027*
Lived in 2–3 homes	0.279	0.449	0.320	0.466	*21,603*	0.315	0.465	0.326	0.469	*22,027*
Lived in 4–9 homes	0.308	0.462	0.304	0.460	*21,603*	0.314	0.464	0.277	0.447	*22,027*
Lived in 10+ homes	0.051	0.219	0.049	0.215	*21,603*	0.061	0.240	0.044	0.206	*22,027*
Father's country of birth/ethnicity										
Australian-born	0.641	0.480	0.653	0.476	*21,317*	0.606	0.489	0.675	0.469	*21,837*
Migrant ESB	0.133	0.340	0.165	0.371	*21,317*	0.146	0.353	0.159	0.365	*21,837*
Migrant Other	0.226	0.418	0.182	0.386	*21,317*	0.249	0.432	0.167	0.373	*21,837*
Mother's country of birth/ethnicity										
Australian-born	0.660	0.474	0.681	0.466	*21,473*	0.625	0.484	0.702	0.458	*21,935*
Migrant ESB	0.127	0.333	0.159	0.366	*21,473*	0.140	0.347	0.148	0.355	*21,935*
Migrant Other	0.213	0.410	0.159	0.366	*21,473*	0.235	0.424	0.150	0.357	*21,935*
Father employed when aged 14										
No	0.036	0.186	0.025	0.157	*21,617*	0.039	0.193	0.032	0.176	*22,042*
Yes	0.868	0.338	0.899	0.301	*21,617*	0.867	0.340	0.905	0.293	*22,042*
Unknown	0.096	0.294	0.076	0.264	*21,617*	0.095	0.293	0.063	0.243	*22,042*

Table A5.1.2 (continued)

Variable	Males					Females				
	Over-educated		Well-matched		*N*	Over-educated		Well-matched		*N*
	mean	s.d	mean	s.d		mean	s.d	mean	s.d	
Father unemployed 6+ months										
No	0.704	0.457	0.802	0.399	*21,617*	0.733	0.442	0.793	0.405	*22,042*
Yes	0.114	0.318	0.085	0.279	*21,617*	0.117	0.321	0.107	0.309	*22,042*
Unknown	0.182	0.386	0.113	0.316	*21,617*	0.150	0.357	0.100	0.300	*22,042*
Mother employed when aged 14										
No	0.405	0.491	0.430	0.495	*21,617*	0.403	0.491	0.413	0.492	*22,042*
Yes	0.509	0.500	0.512	0.500	*21,617*	0.542	0.498	0.548	0.498	*22,042*
Unknown	0.085	0.279	0.058	0.234	*21,617*	0.055	0.227	0.040	0.195	*22,042*
No. siblings	2.636	1.975	2.630	1.951	*21,484*	2.783	2.042	2.703	1.888	*21,931*
Eldest child	0.349	0.477	0.330	0.470	*21,481*	0.338	0.473	0.320	0.467	*21,934*
Father had post-school qualification	0.574	0.495	0.524	0.499	*19,293*	0.562	0.496	0.525	0.499	*19,866*
Mother had post-school qualification	0.339	0.474	0.302	0.459	*19,301*	0.388	0.487	0.358	0.479	*19,861*
Personality measures										
Openness	4.40	1.01	4.26	1.01	*17,022*	4.25	1.05	4.21	1.05	*18,067*
Conscientiousness	4.89	1.01	5.01	0.99	*17,023*	5.18	1.04	5.27	0.99	*18,068*
Extroversion	4.25	1.03	4.28	1.01	*17,026*	4.52	1.14	4.62	1.11	*18,074*
Agreeableness	5.22	0.90	5.10	0.92	*17,025*	5.58	0.84	5.64	0.78	*18,074*
Emotional stability	5.01	1.07	5.08	1.06	*17,023*	5.05	1.09	5.19	1.05	*18,068*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: 's.d' columns report standard deviations; 'N' columns report total number of individuals with non-missing values for each variable.

Appendix 5.2: Full results from regression estimators

Table A5.2.1: ORU earnings functions—Pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)		(II)		(I)		(II)	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Over-educated	-0.151**	0.012	-0.151**	0.015	-0.163**	0.010	-0.167**	0.012
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	-0.066*	0.033	-0.073*	0.036	-0.059*	0.027	-0.073*	0.029
Bachelor Degree	-0.107**	0.030	-0.097**	0.034	-0.093**	0.025	-0.097**	0.028
Advanced Diploma	-0.261**	0.045	-0.257**	0.050	-0.190**	0.033	-0.201**	0.037
Diploma	-0.227**	0.033	-0.212**	0.038	-0.238**	0.029	-0.226**	0.032
Certificate IV	-0.308**	0.035	-0.292**	0.042	-0.288**	0.033	-0.282**	0.037
Certificate III	-0.408**	0.030	-0.413**	0.034	-0.401**	0.027	-0.415**	0.030
Certificate II	-0.500**	0.059	-0.526**	0.074	-0.450**	0.036	-0.457**	0.039
Certificate I	-0.803**	0.185	-0.876**	0.262	-0.528**	0.066	-0.518**	0.080
Year 12	-0.387**	0.033	-0.389**	0.038	-0.346**	0.028	-0.345**	0.031
Year 11	-0.436**	0.034	-0.435**	0.040	-0.384**	0.030	-0.389**	0.033
Year 10 or below	-0.564**	0.035	-0.549**	0.040	-0.493**	0.029	-0.490**	0.032
Multiple qualifications	-0.024	0.013	-0.032*	0.015	0.001	0.010	0.004	0.011
Experience	0.022**	0.006	0.022**	0.007	0.016**	0.005	0.011	0.006
Experience-squared	-0.001*	0.000	-0.001*	0.000	0.000	0.000	0.000	0.000
Experience-cubed	0.000*	0.000	0.000*	0.000	0.000	0.000	0.000	0.000
Occupation tenure	0.007**	0.002	0.007**	0.002	0.006**	0.001	0.006**	0.002
Occupation tenure-squared	0.000*	0.000	0.000*	0.000	0.000	0.000	0.000	0.000
Employer tenure	0.003	0.002	0.004	0.003	0.007**	0.002	0.007**	0.002
Employer tenure-squared	0.000	0.000	0.000	0.000	0.000**	0.000	0.000**	0.000
Years unemployed	-0.023**	0.005	-0.023**	0.006	-0.011*	0.005	-0.013*	0.005
English proficiency (base: English only)								
Good	-0.068**	0.022	-0.090**	0.028	-0.016	0.018	-0.017	0.023
Poor	-0.254**	0.049	-0.312**	0.089	-0.251**	0.048	-0.243**	0.053
L-T health condition	-0.068**	0.015	-0.039**	0.015	-0.036**	0.011	-0.028*	0.012
General health rating			0.000	0.000			0.000	0.000
Mental health rating			0.001*	0.000			0.000	0.000
No. jobs last year	0.016*	0.007	0.019*	0.009	0.008	0.006	0.002	0.007
% time employed last year	0.001**	0.000	0.001	0.000	0.001**	0.000	0.001*	0.000
% time unemployed last year	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.001
Age (base: 15–19 years)								
20–24 years	0.126**	0.026	0.106**	0.030	0.193**	0.022	0.210**	0.027
25–29 years	0.143**	0.035	0.132**	0.042	0.231**	0.030	0.269**	0.035
30–34 years	0.148**	0.043	0.142**	0.052	0.249**	0.037	0.297**	0.043
35–39 years	0.156**	0.048	0.131*	0.060	0.246**	0.043	0.306**	0.049
40–44 years	0.188**	0.053	0.163*	0.065	0.233**	0.047	0.295**	0.053
45–49 years	0.144*	0.059	0.116	0.072	0.213**	0.051	0.282**	0.057
50–54 years	0.125	0.066	0.083	0.081	0.198**	0.057	0.265**	0.064
55–59 years	0.091	0.074	0.035	0.090	0.194**	0.059	0.254**	0.067
60–64 years	-0.026	0.087	-0.074	0.104	0.155*	0.064	0.208**	0.072
Ethnicity (base: Australian-born)								
ATSI	0.014	0.036	0.052	0.044	0.071*	0.028	0.089**	0.033
Migrant ESB	0.073*	0.032	0.070	0.037	0.034	0.022	0.051*	0.025
Migrant Other	-0.045	0.027	-0.044	0.031	0.021	0.021	0.040	0.025

Table A5.2.1 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)		(II)		(I)		(II)	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Marital status (base: Legally married)								
Defacto	-0.039**	0.015	-0.030	0.018	0.002	0.012	0.008	0.014
Separated	-0.078*	0.031	-0.071*	0.033	0.008	0.017	0.013	0.020
Divorced	-0.053	0.031	-0.047	0.035	0.012	0.017	0.014	0.020
Widowed	-0.025	0.054	-0.049	0.057	0.045	0.026	0.052	0.028
Single	-0.123**	0.020	-0.112**	0.021	-0.050**	0.016	-0.032	0.019
No. children had	-0.006	0.015	0.003	0.017	-0.023	0.013	-0.030*	0.014
Resident children by age								
No. aged 0–4 years	0.028	0.017	0.018	0.020	0.055*	0.016	0.062**	0.018
No. aged 5–14 years	0.012	0.016	0.006	0.018	0.011	0.014	0.015	0.015
No. aged 15–24 years	0.016	0.017	0.008	0.020	0.015	0.015	0.023	0.016
No. aged 25+ years	0.042	0.032	0.051	0.039	0.013	0.023	0.046	0.026
Non-resident children by age								
No. aged 0–4 years	0.076**	0.026	0.069*	0.032	0.016	0.122	0.035	0.141
No. aged 5–14 years	0.002	0.018	-0.008	0.023	-0.026	0.025	-0.058	0.037
No. aged 15–24 years	-0.005	0.017	0.000	0.020	0.018	0.015	0.025	0.017
No. aged 25+ years	0.010	0.018	0.004	0.019	0.018	0.014	0.025	0.016
Rest HH income	0.000	0.000	0.000	0.000	0.000**	0.000	0.000**	0.000
Rest HH income-squared	0.000**	0.000	0.000**	0.000	0.000	0.000	0.000	0.000
State/Territory of residence (base: NSW)								
Victoria	-0.038**	0.015	-0.040*	0.018	-0.068**	0.012	-0.062**	0.014
Queensland	-0.063**	0.015	-0.068**	0.018	-0.089**	0.012	-0.089**	0.014
South Australia	-0.132**	0.023	-0.134**	0.028	-0.101**	0.019	-0.106**	0.023
Western Australia	-0.031	0.023	-0.029	0.025	-0.071**	0.017	-0.081**	0.020
Tasmania	-0.054	0.030	-0.057	0.031	-0.063*	0.025	-0.049	0.027
NT	0.074	0.064	0.025	0.078	0.039	0.042	0.022	0.048
ACT	0.043	0.034	0.051	0.037	0.007	0.029	-0.018	0.034
Remoteness of area of residence (base: Major city)								
Inner regional	-0.095**	0.014	-0.093**	0.016	-0.052**	0.011	-0.053**	0.012
Outer regional	-0.102**	0.021	-0.087**	0.026	-0.060**	0.018	-0.044*	0.021
Remote (very remote)	0.042	0.038	0.046	0.039	-0.032	0.030	-0.049	0.032
No. homes lived in (base: Lived in 1 home)								
Lived in 2–3 homes	0.022	0.014	0.025	0.016	0.000	0.011	-0.003	0.012
Lived in 4–9 homes	0.034*	0.015	0.028	0.017	0.026*	0.012	0.014	0.014
Lived in 10+ homes	0.035	0.026	0.039	0.033	0.031	0.019	0.032	0.022
Father's ethnicity (base: Australian-born)								
Migrant ESB	-0.019	0.025	-0.022	0.025	0.019	0.016	0.010	0.019
Migrant Other	0.007	0.021	0.005	0.024	0.029	0.019	0.029	0.022
Mother's ethnicity (base: Australian-born)								
Migrant ESB	-0.037	0.028	-0.041	0.032	-0.013	0.018	-0.023	0.020
Migrant Other	0.014	0.024	0.025	0.028	-0.038	0.021	-0.045	0.024
Father employed when aged 14 (base: No)								
Yes	0.040	0.026	0.020	0.032	0.027	0.023	0.019	0.026
Unknown	0.009	0.034	-0.007	0.041	0.006	0.033	-0.009	0.039
Father unemployed for 6+ months (base: No)								
Yes	-0.016	0.019	-0.012	0.023	0.014	0.014	0.014	0.016
Unknown	-0.016	0.019	-0.034	0.022	0.023	0.018	0.037	0.020
Mother employed when aged 14 (base: No)								
Yes	-0.002	0.012	-0.001	0.014	0.009	0.010	0.006	0.011
Unknown	0.026	0.022	0.026	0.028	-0.046	0.027	-0.045	0.031
No. siblings	-0.002	0.003	0.001	0.004	0.001	0.002	0.002	0.003

Table A5.2.1 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)		(II)		(I)		(II)	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Eldest child	0.005	0.012	0.004	0.014	-0.018*	0.009	-0.025*	0.011
Father had post-school qualification			0.012	0.014			0.022**	0.010
Mother had post-school qualification			-0.006	0.015			-0.002	0.011
Personality measures								
Openness			0.010	0.008			-0.002	0.005
Conscientiousness			0.022**	0.007			0.020**	0.005
Extroversion			-0.006	0.006			0.012**	0.005
Agreeableness			-0.034**	0.008			-0.016**	0.006
Emotional stability			-0.002	0.007			-0.004	0.005
Year controls (base: 2001)								
2002	0.004	0.009	0.017	0.010	0.004	0.009	0.005	0.011
2003	0.032**	0.009	0.032**	0.010	0.012	0.010	0.017	0.011
2004	0.053**	0.009	0.055**	0.011	0.014	0.010	0.016	0.011
2005	0.062**	0.010	0.072**	0.011	0.030**	0.011	0.036**	0.012
2006	0.059**	0.010	0.072**	0.012	0.047**	0.010	0.044**	0.011
2007	0.085**	0.010	0.091**	0.012	0.062**	0.010	0.066**	0.012
2008	0.074**	0.010	0.084**	0.013	0.042**	0.011	0.045**	0.012
Intercept term	2.831**	0.067	2.838**	0.094	2.644**	0.055	2.566**	0.082
N	20,288		15,146		20,806		16,357	
(No. individuals)	(5,231)		(3,260)		(5,370)		(3,637)	
R-squared	0.3051		0.3065		0.2924		0.2960	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Table A5.2.2: ORU earnings functions—Fixed effects and first-differences estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	Fixed effects		First-differences		Fixed effects		First-differences	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Over-educated	-0.036*	0.015	-0.043*	0.020	-0.060**	0.014	-0.065**	0.018
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	-0.031	0.041	0.019	0.062	-0.054	0.037	-0.042	0.064
Bachelor Degree	-0.037	0.033	0.055	0.057	-0.066	0.035	-0.064	0.055
Advanced Diploma	-0.052	0.082	0.142	0.080	-0.163*	0.074	-0.111	0.095
Diploma	-0.073	0.053	0.073	0.073	-0.112*	0.050	-0.002	0.081
Certificate IV	0.036	0.061	0.083	0.080	-0.104*	0.051	-0.009	0.083
Certificate III	0.004	0.058	0.085	0.077	-0.119*	0.053	-0.027	0.090
Certificate II	-0.146	0.092	-0.079	0.125	-0.177*	0.073	-0.066	0.122
Certificate I	-0.109	0.091	0.035	0.144	0.026	0.146	0.070	0.162
Year 12	-0.082	0.060	0.002	0.079	-0.131*	0.054	-0.039	0.086
Year 11	-0.039	0.063	0.063	0.089	-0.232**	0.077	-0.120	0.120
Year 10 or below	-0.089	0.075	-0.017	0.088	-0.176**	0.064	-0.111	0.101
Multiple qualifications	-0.005	0.018	0.027	0.024	0.018	0.019	0.028	0.029
Experience	0.058**	0.019	0.021	0.037	0.055**	0.016	0.099*	0.044
Experience-squared	-0.002**	0.000	-0.002**	0.000	-0.001**	0.000	-0.002**	0.001
Experience-cubed	0.000**	0.000	0.000**	0.000	0.000*	0.000	0.000**	0.000
Occupation tenure	0.001	0.001	0.000	0.001	0.003*	0.001	0.001	0.002
Occupation tenure-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Employer tenure	0.000	0.002	0.001	0.002	0.000	0.002	-0.002	0.002
Employer tenure-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Years unemployed	-0.045	0.036	-0.023	0.098	-0.045	0.040	0.085	0.162
English proficiency (base: English only)								
Good	-0.023	0.018	-0.002	0.021	0.010	0.023	0.014	0.020
Poor	-0.068	0.051	-0.073	0.050	-0.082	0.067	-0.183**	0.071
L-T health condition	-0.005	0.007	0.001	0.008	0.014	0.009	0.007	0.009
No. jobs last year	0.020**	0.006	0.006	0.006	0.009	0.005	0.006	0.005
% time employed last year	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% time unemployed last year	0.001	0.001	0.000	0.001	0.000	0.000	-0.001	0.001
Age (base: 15–19 years)								
20–24 years	0.104**	0.031	0.027	0.039	0.161**	0.026	0.064	0.034
25–29 years	0.089*	0.037	0.011	0.045	0.168**	0.034	0.043	0.042
30–34 years	0.077	0.042	0.017	0.050	0.168**	0.042	0.056	0.050
35–39 years	0.073	0.046	0.011	0.053	0.136**	0.047	0.040	0.056
40–44 years	0.104*	0.048	0.033	0.055	0.115*	0.051	-0.003	0.061
45–49 years	0.112*	0.050	0.044	0.057	0.100	0.054	0.004	0.065
50–54 years	0.116*	0.053	0.039	0.060	0.103	0.058	0.025	0.069
55–59 years	0.103	0.057	0.034	0.064	0.080	0.064	0.005	0.074
60–64 years	0.154*	0.073	0.066	0.080	0.074	0.072	0.030	0.083
Marital status (base: Legally married)								
Defacto	-0.023	0.014	-0.008	0.017	0.006	0.014	0.013	0.019
Separated	-0.008	0.020	0.005	0.024	0.001	0.018	0.023	0.024
Divorced	-0.016	0.029	-0.005	0.028	0.012	0.023	0.047	0.031
Widowed	0.117	0.130	0.052	0.078	0.001	0.045	0.016	0.055
Single	-0.053**	0.019	-0.034	0.026	-0.012	0.020	-0.018	0.026
No. children had	-0.010	0.012	-0.004	0.018	-0.011	0.018	0.012	0.026
Resident children by age								
No. aged 0–4 years	0.008	0.012	-0.002	0.015	0.024	0.020	0.038	0.025
No. aged 5–14 years	-0.002	0.011	-0.004	0.014	-0.009	0.020	0.033	0.021
No. aged 15–24 years	0.016	0.012	0.015	0.015	-0.017	0.020	0.017	0.021
No. aged 25+ years	0.048*	0.023	-0.005	0.027	-0.034	0.026	0.028	0.028

Table A5.2.2 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	Fixed effects		First-differences		Fixed effects		First-differences	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Non-resident children by age								
No. aged 0–4 years	0.014	0.023	0.024	0.029	0.024	0.050	-0.014	0.059
No. aged 5–14 years	-0.006	0.015	-0.016	0.017	-0.010	0.032	-0.013	0.027
No. aged 15–24 years	0.008	0.012	0.006	0.018	-0.019	0.021	0.021	0.021
No. aged 25+ years	0.030*	0.014	0.007	0.016	-0.023	0.022	0.025	0.021
Rest HH income	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rest HH income-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State/Territory of residence (base: NSW)								
Victoria	0.026	0.051	0.056	0.058	0.027	0.045	0.019	0.062
Queensland	-0.089**	0.033	-0.073	0.049	0.040	0.042	0.064	0.056
South Australia	-0.033	0.056	-0.046	0.082	0.028	0.054	0.072	0.071
Western Australia	-0.009	0.050	0.044	0.067	-0.011	0.075	-0.019	0.075
Tasmania	-0.161	0.097	-0.221	0.183	-0.076	0.081	-0.005	0.092
NT	0.080	0.073	0.129	0.092	0.070	0.102	0.037	0.126
ACT	0.024	0.038	0.039	0.042	0.012	0.064	0.022	0.068
Remoteness of area of residence (base: Major city)								
Inner regional	-0.024	0.022	-0.026	0.024	-0.062**	0.023	-0.028	0.028
Outer regional	0.014	0.029	0.024	0.037	-0.085*	0.036	-0.003	0.043
Remote (very remote)	0.035	0.054	0.075	0.069	-0.062	0.068	-0.006	0.079
No. homes lived in (base: Lived in 1 home)								
Lived in 2–3 homes	-0.011	0.026	0.008	0.024	-0.027	0.027	-0.002	0.030
Lived in 4–9 homes	0.014	0.023	-0.016	0.018	0.040*	0.019	0.052*	0.022
Lived in 10+ homes	0.007	0.039	-0.005	0.037	-0.002	0.043	-0.007	0.049
Year controls (base: 2001)								
2002	0.012	0.018	0.060	0.035	-0.017	0.017	-0.040	0.044
2003	0.041	0.033	0.127	0.068	-0.020	0.029	-0.066	0.085
2004	0.065	0.049	0.191	0.102	-0.033	0.041	-0.101	0.126
2005	0.082	0.065	0.243	0.136	-0.035	0.055	-0.125	0.168
2006	0.091	0.081	0.292	0.169	-0.030	0.068	-0.146	0.209
2007	0.121	0.097	0.363	0.203	-0.022	0.081	-0.161	0.250
2008	0.118	0.112	0.400	0.236	-0.051	0.094	-0.217	0.291
Intercept term	2.487**	0.301			2.266**	0.220		
N	20,636		13,959		21,035		13,981	
(No. individuals)	(5,323)		(3,659)		(5,449)		(3,765)	
Within R-squared	0.0674				0.0421			
Between R-squared	0.2060				0.1591			
Overall R-squared	0.1616		0.0154		0.1136		0.0130	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Appendix 5.3: Details for cross-sectional matching estimators

Table A5.3.1: Logit models for probability over-educated—Propensity score model results

Dependent variable: Over-educated	Males		Females	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Coeff.*	*SE*
Highest education (base: Postgraduate Degree)				
Graduate Diploma or Certificate	0.338**	0.122	0.581**	0.127
Bachelor Degree	0.593**	0.106	1.007**	0.117
Advanced Diploma	1.439**	0.144	1.473**	0.144
Diploma	1.485**	0.118	2.596**	0.129
Certificate IV	0.899**	0.132	2.388**	0.140
Certificate III	-0.440**	0.116	0.089	0.129
Certificate II	-0.143	0.209	1.021**	0.185
Year 12	-0.068	0.121	0.472**	0.126
Year 11	0.307*	0.132	0.822**	0.135
Multiple qualifications	-0.147**	0.049	-0.141**	0.045
Experience	-0.080**	0.025	-0.062**	0.023
Experience-squared	0.005**	0.001	0.002	0.001
Experience-cubed	0.000**	0.000	0.000	0.000
Occupation tenure	-0.128**	0.008	-0.099**	0.008
Occupation tenure-squared	0.002**	0.000	0.001**	0.000
Employer tenure	-0.017	0.009	0.016	0.011
Employer tenure-squared	0.000	0.000	-0.002**	0.000
Years unemployed	0.190**	0.018	0.113	0.073
L-T health condition	0.282**	0.060	0.282**	0.061
English proficiency (base: English only)				
Good	0.189*	0.093	0.111	0.091
Poor	1.159**	0.255	0.872**	0.267
No. jobs last year	0.064*	0.031	0.049	0.029
% time employed last year	-0.004*	0.002	-0.005**	0.001
% time unemployed last year	0.000	0.003	0.001	0.002
Age (base: 15–19 years)				
20–24 years	-0.887**	0.126	-0.515**	0.119
25–29 years	-0.952**	0.154	-0.590**	0.139
30–34 years	-0.811**	0.186	-0.351*	0.163
35–39 years	-0.786**	0.210	-0.218	0.177
40–44 years	-0.900**	0.233	-0.281	0.408
45–49 years	-0.948**	0.254	-0.115	0.199
50–54 years	-1.120**	0.277	0.012	0.213
55–59 years	-1.240**	0.305	-0.046	0.231
60–64 years	-1.096**	0.352	0.155	0.275
Ethnicity (base: Australian-born)				
ATSI	-0.066	0.180	-0.085	0.170
Migrant ESB	-0.036	0.109	0.216*	0.101
Migrant Other	0.467**	0.099	0.748**	0.091
Marital status (base: Legally married)				
Defacto	0.279**	0.067	0.163**	0.060
Separated	0.655**	0.139	-0.084	0.110
Divorced	0.864**	0.119	-0.113	0.085
Widowed	-2.170*	1.029	0.055	0.196
Single	0.418**	0.070	0.306**	0.065

Table A5.3.1 (continued)

Dependent variable: Over-educated	Males		Females	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Coeff.*	*SE*
No. children had	0.077	0.081	0.249**	0.082
Resident children by age				
No. aged 0–4 years	-0.004	0.093	-0.104	0.094
No. aged 5–14 years	-0.188*	0.086	-0.240**	0.087
No. aged 15–24 years	-0.115	0.091	-0.372**	0.091
No. aged 25+ years	0.151	0.161	0.021	0.166
Non-resident children by age				
No. aged 0–4 years	-0.085	0.163	0.137	0.590
No. aged 5–14 years	-0.091	0.101	-0.023	0.179
No. aged 15–24 years	-0.224*	0.091	-0.353**	0.098
No. aged 25+ years	-0.045	0.092	-0.153	0.092
Rest HH income	0.000*	0.000	0.000	0.000
Rest HH income-squared	0.000	0.000	0.000	0.000
State/Territory of residence (base: NSW)				
Victoria	-0.086	0.056	0.122*	0.052
Queensland	0.130*	0.060	0.060	0.061
South Australia	0.049	0.077	0.101	0.073
Western Australia	0.178*	0.078	-0.326	0.284
Tasmania	0.030	0.137	0.230	0.125
NT	0.231	0.211	-0.400*	0.199
ACT	-0.336*	0.141	-0.452**	0.139
Remoteness of area of residence (base: Major city)				
Inner regional	0.365**	0.055	-0.035	0.057
Outer regional	0.383**	0.072	0.228**	0.076
Remote (very remote)	-0.338*	0.165	0.028	0.152
No. homes lived in (base: Lived in 1 home)				
Lived in 2–3 homes	-0.152**	0.053	0.092	0.050
Lived in 4–9 homes	-0.191**	0.073	0.027	0.052
Lived in 10+ homes	-0.274	0.149	-0.596	0.316
Father's ethnicity (base: Australian-born)				
Migrant ESB	-0.140	0.081	-0.178*	0.078
Migrant Other	-0.160	0.088	-0.059	0.082
Mother's ethnicity (base: Australian-born)				
Migrant ESB	-0.196*	0.085	0.029	0.079
Migrant Other	0.062	0.094	0.083	0.091
Father employed when aged 14 (base: No)				
Yes	-0.285*	0.125	-0.112	0.109
Unknown	-0.399**	0.149	0.408**	0.133
Father unemployed for 6+ months (base: No)				
Yes	0.194**	0.070	0.008	0.068
Unknown	0.281**	0.075	0.002	0.076
Mother employed when aged 14 (base: No)				
Yes	0.002	0.055	-0.099*	0.041
Unknown	-0.036	0.095	-0.181	0.110
No. siblings	0.057**	0.011	0.038**	0.011
Eldest child	0.058	0.043	0.024	0.041
Year controls (base: 2001)				
2002	0.073	0.081	-0.052	0.078
2003	0.040	0.080	-0.029	0.077
2004	0.012	0.082	0.002	0.077
2005	-0.039	0.082	0.009	0.076
2006	0.051	0.081	0.094	0.076

Table A5.3.1 (continued)

Dependent variable: Over-educated	Males		Females	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Coeff.*	*SE*
2007	0.011	0.082	0.033	0.076
2008	0.052	0.081	0.050	0.076
Intercept term	0.391	0.273	-0.489*	0.247
N	17,270		17,214	
Pseudo R-squared	0.1693		0.1734	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Models also contain various interaction terms, based on interacting the following: *age* and *number of children*; *age* and *experience*; *age* and *divorced*; *age* and *English proficiency*; *ethnicity* and *number of homes lived in during past 10 years*; *years unemployed* and *age*, *State/Territory of residence*, *remoteness of area of residence*, *number of homes lived in during past 10 years*; and, *proportion time employed last year* and *age*, *State/Territory of residence*. Full sets of results are available from the author on request.

Table A5.3.2: Balancing tests results—Differences in mean characteristics of over-educated and well-matched individuals pre- and post-matching (using nearest neighbour and kernel matching weights)

Individual characteristics	Males			Females		
	Pre-matching diff.s	Post-matching diff.s (NN weights)	Post-matching diff.s (Kernel weights)	Pre-matching diff.s	Post-matching diff.s (NN weights)	Post-matching diff.s (Kernel weights)
Graduate Diploma or Certificate	-0.006	0.000	0.000	-0.034**	0.000	0.000
Bachelor Degree	0.077**	0.000	0.000	0.054**	0.000	0.000
Advanced Diploma	0.029**	0.000	0.000	0.012**	0.000	0.000
Diploma	0.090**	0.000	0.000	0.138**	0.000	0.000
Certificate IV	0.041**	0.000	0.000	0.072**	0.000	0.000
Certificate III	-0.113**	0.000	0.000	-0.038**	0.000	0.000
Certificate II	0.004*	0.000	0.000	0.007**	0.000	0.000
Year 12	0.058**	0.000	0.000	0.025**	0.000	0.000
Year 11	0.035**	0.000	0.000	0.022**	0.000	0.000
Multiple qualifications	0.047**	0.018	0.010	0.058**	0.003	0.004
Experience	-4.503**	-0.281	-0.327	-3.898**	0.181	0.038
Experience-squared	-156.22**	-11.57	-12.76	-132.49**	6.168	2.136
Experience-cubed	-5396.04**	-481.54	-526.03	-4303.88**	205.85	89.62
Occupation tenure	-5.458**	0.045	-0.074	-4.734**	-0.073	-0.129
Occupation tenure-squared	-135.42**	0.854	-0.820	-120.81**	1.095	-0.493
Employer tenure	-2.983**	0.013	-0.139	-2.868**	0.056	-0.044
Employer tenure-squared	-62.71**	1.545	-2.878	-62.26**	1.908	0.449
Years unemployed	0.263**	-0.027	-0.024	0.169**	0.007	0.035
L-T health condition	0.010	-0.005	-0.001	0.008	0.004	-0.001
Good English proficiency	0.042**	0.005	0.003	0.063**	-0.001	-0.005
Poor English proficiency	0.006**	0.004	0.003*	0.007**	0.001	0.002
No. jobs last year	0.112**	0.020	-0.001	0.088**	-0.002	-0.017
% time employed last year	-5.534**	0.208	-0.590	-5.218**	-0.245	-0.904*
% time unemployed last year	2.777**	-0.467	0.160	2.067**	-0.376	0.294
20–24 years	0.068**	-0.009	0.003	0.066**	-0.014*	-0.005
25–29 years	0.026**	0.008	0.003	0.024**	0.000	-0.005
30–34 years	-0.004	-0.004	0.004	0.016**	0.000	-0.004
35–39 years	-0.026**	0.004	-0.003	0.005	-0.007	-0.002
40–44 years	-0.036**	0.005	-0.007	-0.030**	0.014*	0.014**
45–49 years	-0.031**	-0.002	0.001	-0.038**	0.005	0.006
50–54 years	-0.025**	-0.008	0.000	-0.040**	-0.005	-0.005
55–59 years	-0.020**	0.001	-0.001	-0.025**	0.005	-0.002
60–64 years	-0.006	-0.002	-0.004	-0.014**	-0.001	0.000
ATSI	0.001	0.000	0.002	0.003	0.003	0.003
Migrant ESB	-0.031**	0.001	-0.003	-0.005	-0.001	-0.001
Migrant Other	0.058**	-0.008	0.002	0.082**	-0.005	0.000
Defacto	0.002	-0.003	-0.007	0.034**	-0.008	-0.002
Separated	0.005	-0.003	-0.004	-0.005	0.004	0.002
Divorced	0.008*	0.014**	0.009**	-0.015**	0.006	0.000
Widowed	-0.003**	0.000	0.000	-0.008**	-0.004	-0.002
Single	0.138**	0.010	0.006	0.092**	-0.002	0.001
No. children had	-0.381**	-0.010	0.008	-0.297**	0.007	0.006
No. resident children 0–4 years	-0.039**	-0.017	-0.007	0.028**	-0.015	-0.005
No. resident children 5–14 years	-0.161**	0.012	0.006	-0.040**	0.022	0.011
No. resident children 15–24 years	-0.068**	0.002	0.005	-0.102**	-0.009	-0.001
No. resident children 25+ years	-0.004	-0.003	0.000	-0.009**	-0.004	-0.003
No. non-resident children 0–4 years	0.001	0.005	0.001	0.001	0.000	0.000

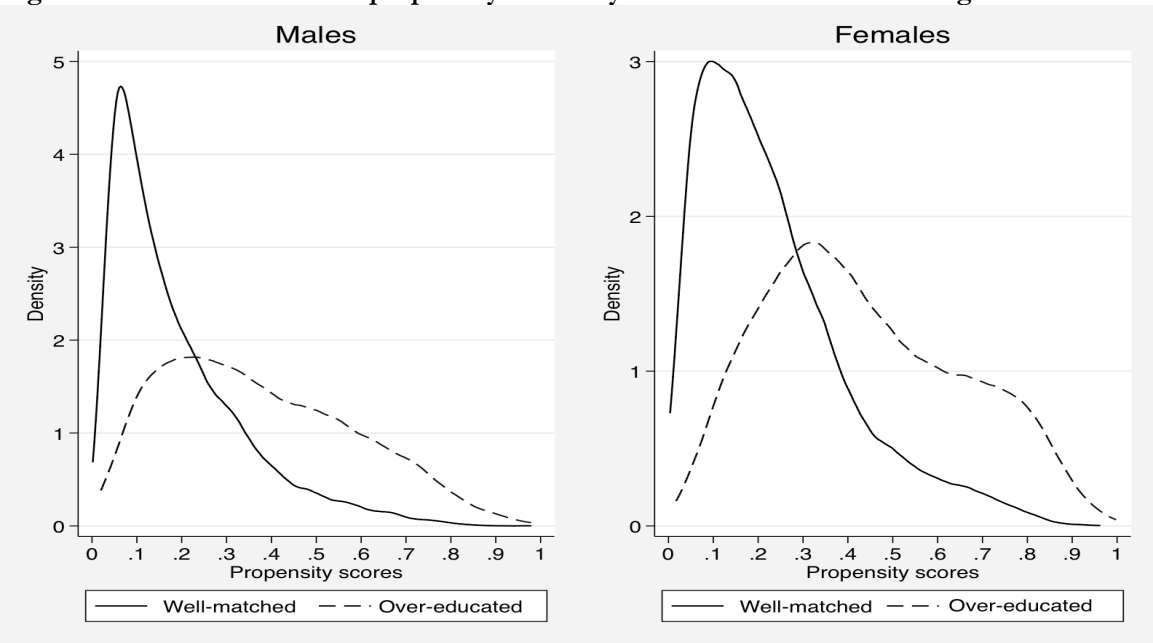
Table A5.3.2 (continued)

Individual characteristics	Males			Females		
	Pre-matching diff.s	Post-matching diff.s (NN weights)	Post-matching diff.s (Kernel weights)	Pre-matching diff.s	Post-matching diff.s (NN weights)	Post-matching diff.s (Kernel weights)
No. non-resident children 5–14 years	-0.006	0.008	0.002	0.003	0.005*	0.002
No. non-resident children 15–24 years	-0.042**	-0.003	0.010	-0.048**	0.005	0.005
No. non-resident children 25+ years	-0.059**	-0.002	-0.005	-0.130**	0.003	-0.006
Rest HH income	4909.46**	1175.73	-24.13	973.85	-301.57	62.54
Rest HH income-squared	$8.04x10^{8}$**	$1.68x10^{8}$	$-4.75x10^{7}$	$2.68x10^{8}$	$-4.17x10^{8}$	$-9.78x10^{7}$
Victoria	-0.009	0.005	0.015*	0.030**	0.006	0.000
Queensland	0.008	0.002	-0.007	-0.005	0.002	0.000
South Australia	0.006	0.000	0.006	0.002	-0.005	-0.001
Western Australia	0.006	-0.003	-0.004	0.003	-0.001	-0.006
Tasmania	-0.003	-0.003	0.000	-0.004	0.007*	0.004
NT	0.003*	0.000	0.001	0.004**	0.002	0.001
ACT	-0.004	0.000	0.000	-0.005*	0.000	0.000
Inner regional	-0.017*	-0.001	0.005	-0.038**	0.009	0.004
Outer regional	0.017**	-0.002	0.003	0.009	0.004	0.007
Remote (very remote)	-0.003	0.002	0.001	-0.001	-0.001	-0.001
Lived in 2–3 homes	-0.041**	0.006	-0.001	-0.010	0.008	0.004
Lived in 4–9 homes	0.004	-0.002	0.005	0.038**	-0.004	0.005
Lived in 10+ homes	0.002	-0.003	-0.006	0.017**	0.002	-0.005
Father migrant ESB	-0.032**	0.004	-0.004	-0.013*	-0.013	0.001
Father migrant Other	0.044**	0.002	0.000	0.082**	0.008	-0.005
Mother migrant ESB	-0.033**	-0.011	-0.007	-0.008	0.010	0.002
Mother migrant Other	0.054**	0.013	0.006	0.085**	-0.001	-0.001
Father employed when aged 14	-0.031**	0.001	-0.003	-0.038**	0.011	0.001
Father's employment unknown	0.020**	-0.005	0.000	0.032**	-0.012*	-0.006
Father unemployed for 6+ months	0.029**	0.012	0.005	0.009	0.006	0.007
Father's unemployment unknown	0.069**	-0.004	0.003	0.050**	-0.010	-0.004
Mother employed when aged 14	-0.003	0.022	0.006	-0.006	-0.011	-0.003
Mother's employment unknown	0.027**	0.000	-0.001	0.015**	-0.004	-0.006
No. siblings	0.007	0.026	0.034	0.080**	-0.030	0.023
Eldest child	0.019**	-0.025*	-0.006	0.018*	0.006	-0.001
2002	-0.004	0.002	0.004	-0.013*	-0.008	-0.003
2003	0.002	0.006	0.003	-0.007	0.003	0.002
2004	0.000	0.003	-0.001	-0.003	0.008	0.003
2005	-0.003	-0.006	-0.003	0.007	-0.004	-0.002
2006	0.007	-0.012	-0.002	0.012*	0.000	0.003
2007	0.006	0.000	-0.004	0.008	-0.008	-0.001
2008	0.013*	-0.004	-0.002	0.007	0.002	0.000
Smith and Todd (2005) Hotelling T-test for joint significance (p-values)		0.3668	0.8834		0.4960	0.9583
N	21,617	6,624	17,052	22,042	8,030	16,915
(No. over-educated)	(4,302)	(3,970)	(3,946)	(5,362)	(5,023)	(5,000)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Reported differences are mean(over-educated) – mean(well-matched).
** and * indicate statistically significant differences at the 1% and 5% levels; for 'Post-matching diff.s' columns, these are the results from the Smith and Todd (2005) balancing tests.
Bold figures indicate differences exceed one-quarter of the standard deviation of the variable in the pre-matching sample; these are the results from the Ho et al. (2007) balancing tests. (No such differences actually exist.)
Results from the Dehejia and Wahba (2002) balancing tests are not presented, but are available from the author on request.

Figure A5.3.1: Distribution of propensity scores by over-education status and gender



SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

Table A5.3.3: Samples pre- and post-matching by highest education level and gender

	Pre-matching		Post-matching			
			Nearest neighbour matching		*Kernel matching*	
	Over-educ.	W-M	Over-educ.	W-M	Over-educ.	W-M
Males						
Postgraduate Degree	154	987	152	122	152	943
Graduate Diploma or Cert.	235	1,054	221	179	221	1,018
Bachelor Degree	972	2,577	898	627	894	2,427
Advanced Diploma	199	298	182	91	175	285
Diploma	577	757	537	292	534	713
Certificate IV	293	474	278	137	273	453
Certificate III	646	4,553	592	493	591	4,331
Certificate II	60	173	48	35	47	150
Year 12	788	2,168	710	451	709	1,947
Year 11	378	910	352	227	350	839
Total	4,302	13,951	3,970	2,654	3,946	13,106
Females						
Postgraduate Degree	123	748	109	90	109	711
Graduate Diploma or Cert.	353	1,669	340	251	339	1,533
Bachelor Degree	1,329	3,232	1,258	803	1,258	2,957
Advanced Diploma	255	592	243	116	239	544
Diploma	897	491	842	323	842	471
Certificate IV	480	291	445	171	434	272
Certificate III	447	2,018	421	327	419	1,907
Certificate II	97	184	93	59	88	173
Year 12	945	2,526	861	607	861	2,384
Year 11	436	993	411	260	411	963
Total	5,362	12,744	5,023	3,007	5,000	11,915

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTE: 'W-M' refers to well-matched individuals.

Table A5.3.4: Distribution of propensity scores in post-matching samples, and reliance on well-matched individuals to identify nearest neighbour matching estimates across distribution

	Males			Females		
	Over-educ.	W-M	Avg. no. times W-M used	Over-educ.	W-M	Avg. no. times W-M used
$0.00 \leq P(X) < 0.10$	298	281	1.07	179	158	1.12
$0.10 \leq P(X) < 0.20$	683	589	1.15	579	506	1.15
$0.20 \leq P(X) < 0.30$	719	577	1.25	831	648	1.28
$0.30 \leq P(X) < 0.40$	633	465	1.36	894	598	1.49
$0.40 \leq P(X) < 0.50$	518	303	1.74	730	411	1.77
$0.50 \leq P(X) < 0.60$	451	211	2.12	547	286	1.93
$0.60 \leq P(X) < 0.70$	339	140	2.39	496	211	2.36
$0.70 \leq P(X) < 0.80$	222	74	3.23	438	138	3.27
$0.80 \leq P(X) < 0.90$	83	13	6.62	273	47	6.09
$0.90 \leq P(X) \leq 1.00$	24	1	3.00	56	4	6.25
Total	3,970	2,654		5,023	3,007	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).
NOTE: 'W-M' refers to well-matched individuals.

Appendix 5.4: Full results from combined matching and regression estimators

Table A5.4.1: ORU earnings functions—Combined matching and pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*		Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Over-educated	-0.149**	0.014	-0.136**	0.012	-0.176**	0.014	-0.160**	0.011
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	-0.079	0.044	-0.091*	0.039	-0.114**	0.042	-0.089*	0.038
Bachelor Degree	-0.108**	0.041	-0.113**	0.037	-0.170**	0.036	-0.146**	0.033
Advanced Diploma	-0.332**	0.053	-0.312**	0.051	-0.291**	0.048	-0.260**	0.043
Diploma	-0.253**	0.041	-0.248**	0.038	-0.313**	0.043	-0.297**	0.038
Certificate IV	-0.294**	0.045	-0.279**	0.041	-0.343**	0.042	-0.331**	0.039
Certificate III	-0.402**	0.041	-0.405**	0.037	-0.429**	0.039	-0.421**	0.035
Certificate II	-0.488**	0.056	-0.478**	0.051	-0.507**	0.051	-0.452**	0.042
Year 12	-0.362**	0.044	-0.367**	0.040	-0.433**	0.038	-0.399**	0.035
Year 11	-0.413**	0.046	-0.410**	0.042	-0.464**	0.043	-0.430**	0.036
Multiple qualifications	-0.015	0.018	-0.009	0.016	-0.034*	0.016	-0.017	0.013
Experience	0.035**	0.008	0.031**	0.008	0.012	0.007	0.011	0.006
Experience-squared	-0.001**	0.000	-0.001**	0.000	-0.001	0.000	0.000	0.000
Experience-cubed	0.000**	0.000	0.000**	0.000	0.000	0.000	0.000	0.000
Occupation tenure	0.011**	0.003	0.008**	0.002	0.011**	0.003	0.008**	0.002
Occupation tenure-squared	0.000*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Employer tenure	0.006	0.003	0.009**	0.003	0.004	0.004	0.005	0.003
Employer tenure-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Years unemployed	-0.027**	0.005	-0.026**	0.005	-0.013*	0.006	-0.014**	0.005
English proficiency (base: English only)								
Good	-0.085**	0.031	-0.081**	0.027	-0.011	0.030	-0.002	0.025
Poor	-0.375**	0.068	-0.328**	0.070	-0.221**	0.053	-0.196**	0.055
L-T health condition	-0.055**	0.020	-0.046**	0.017	-0.042*	0.018	-0.034*	0.014
No. jobs last year	0.009	0.009	0.010	0.008	0.001	0.009	0.001	0.007
% time employed last year	0.001*	0.001	0.001*	0.000	0.002*	0.001	0.002**	0.001
% time unemployed last year	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001
Age (base: 15–19 years)								
20–24 years	0.135**	0.035	0.127**	0.031	0.163**	0.031	0.177**	0.024
25–29 years	0.124*	0.049	0.138**	0.043	0.215**	0.038	0.236**	0.031
30–34 years	0.118*	0.060	0.111*	0.054	0.250**	0.046	0.281**	0.037
35–39 years	0.092	0.072	0.108	0.064	0.288**	0.056	0.284**	0.044
40–44 years	0.175*	0.077	0.168*	0.070	0.286**	0.056	0.289**	0.046
45–49 years	0.136	0.085	0.152*	0.077	0.261**	0.063	0.266**	0.051
50–54 years	0.107	0.091	0.097	0.084	0.228**	0.066	0.234**	0.055
55–59 years	0.112	0.102	0.081	0.093	0.167*	0.073	0.213**	0.061
60–64 years	-0.072	0.124	-0.083	0.112	0.115	0.094	0.165*	0.072
Ethnicity (base: Australian-born)								
ATSI	0.007	0.048	-0.002	0.045	0.030	0.044	0.050	0.035
Migrant ESB	0.006	0.045	0.037	0.038	0.039	0.032	0.052*	0.026
Migrant Other	-0.044	0.036	-0.043	0.031	0.021	0.040	0.033	0.032

Table A5.4.1 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*		Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Marital status (base: Legally married)								
Defacto	-0.035	0.021	-0.036	0.018	0.007	0.022	-0.009	0.017
Separated	-0.121*	0.047	-0.098*	0.040	0.033	0.027	0.031	0.022
Divorced	-0.040	0.044	-0.048	0.041	0.022	0.030	0.008	0.027
Widowed	0.152	0.085	0.163*	0.072	-0.003	0.063	-0.005	0.054
Single	-0.094**	0.022	-0.095**	0.019	-0.025	0.028	-0.041	0.023
No. children had	-0.006	0.022	0.004	0.019	-0.053**	0.016	-0.040**	0.015
Resident children by age								
No. aged 0–4 years	0.019	0.026	0.022	0.021	0.122**	0.030	0.091**	0.022
No. aged 5–14 years	0.028	0.024	0.022	0.020	0.028	0.019	0.011	0.018
No. aged 15–24 years	-0.007	0.027	0.002	0.023	0.037	0.021	0.018	0.018
No. aged 25+ years	0.021	0.051	0.020	0.048	0.042	0.036	0.011	0.029
Non-resident children by age								
No. aged 0–4 years	0.076	0.047	0.057	0.037	0.161**	0.050	0.147**	0.050
No. aged 5–14 years	0.014	0.029	0.000	0.026	-0.004	0.042	0.002	0.030
No. aged 15–24 years	0.014	0.027	-0.003	0.024	0.037	0.023	0.023	0.018
No. aged 25+ years	-0.033	0.025	-0.039	0.022	0.061**	0.020	0.040*	0.018
Rest HH income	0.000*	0.000	0.000	0.000	0.000**	0.000	0.000**	0.000
Rest HH income-squared	0.000**	0.000	0.000**	0.000	0.000**	0.000	0.000*	0.000
State/Territory of residence (base: NSW)								
Victoria	-0.039	0.020	-0.031	0.018	-0.062**	0.021	-0.048**	0.017
Queensland	-0.095**	0.020	-0.093**	0.019	-0.090**	0.021	-0.069**	0.018
South Australia	-0.074**	0.028	-0.072**	0.023	-0.076**	0.027	-0.065**	0.022
Western Australia	-0.050	0.028	-0.053*	0.025	-0.102**	0.028	-0.070**	0.025
Tasmania	-0.063	0.051	-0.081	0.049	-0.030	0.039	-0.018	0.033
NT	0.126	0.086	0.088	0.081	-0.048	0.054	-0.017	0.044
ACT	0.013	0.051	0.008	0.047	0.066	0.038	0.060	0.035
Remoteness of area (base: Major city)								
Inner regional	-0.090**	0.018	-0.092**	0.016	-0.053**	0.018	-0.049**	0.014
Outer regional	-0.131**	0.025	-0.110**	0.022	-0.032	0.037	-0.050	0.032
Remote (very remote)	0.053	0.047	0.078*	0.037	0.058	0.046	0.031	0.036
No. homes lived in (base: Lived in 1 home)								
Lived in 2–3 homes	0.022	0.018	0.024	0.016	0.018	0.020	0.015	0.017
Lived in 4–9 homes	0.046*	0.019	0.048**	0.017	0.043*	0.021	0.036*	0.018
Lived in 10+ homes	0.014	0.035	0.024	0.030	0.057	0.032	0.039	0.026
Father's ethnicity (base: Australian-born)								
Migrant ESB	0.002	0.033	-0.009	0.024	0.037	0.032	0.026	0.024
Migrant Other	-0.008	0.032	0.004	0.028	0.007	0.027	0.014	0.024
Mother's ethnicity (base: Australian-born)								
Migrant ESB	0.012	0.031	-0.003	0.028	-0.020	0.027	-0.028	0.023
Migrant Other	0.034	0.032	0.008	0.030	-0.018	0.032	-0.045	0.027
Father employed when aged 14 (base: No)								
Yes	0.086*	0.037	0.055	0.036	0.054	0.034	0.035	0.030
Unknown	0.072	0.044	0.041	0.042	0.052	0.044	0.046	0.036
Father unemployed for 6+ months (base: No)								
Yes	-0.030	0.026	-0.047*	0.022	0.027	0.022	0.023	0.019
Unknown	-0.012	0.024	-0.026	0.021	0.026	0.026	0.030	0.019

Table A5.4.1 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*		Post-matching sample: *Nearest neighbour*		Post-matching sample: *Kernel matching*	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Mother employed when aged 14 (base: No)								
Yes	-0.022	0.016	-0.022	0.015	-0.002	0.016	0.000	0.013
Unknown	-0.008	0.029	0.011	0.027	-0.017	0.034	-0.037	0.028
No. siblings	-0.004	0.004	-0.003	0.003	-0.003	0.004	-0.001	0.003
Eldest child	-0.003	0.016	-0.001	0.014	-0.013	0.016	-0.016	0.014
Year controls (base: 2001)								
2002	0.000	0.021	-0.002	0.015	0.002	0.022	-0.005	0.015
2003	0.022	0.022	0.041*	0.017	0.038	0.020	0.018	0.015
2004	0.040	0.021	0.044**	0.015	0.011	0.020	0.018	0.016
2005	0.041	0.022	0.054**	0.017	0.046	0.028	0.038*	0.018
2006	0.031	0.022	0.035*	0.017	0.061**	0.022	0.051**	0.016
2007	0.063**	0.022	0.075**	0.016	0.055*	0.026	0.051*	0.022
2008	0.037	0.023	0.054**	0.017	0.054*	0.022	0.045**	0.016
Intercept term	2.772**	0.094	2.804**	0.077	2.649**	0.086	2.632**	0.072
N	7,868		16,938		9,970		16,776	
(No. individuals)	(2,860)		(4,325)		(3,237)		(4,368)	
R-squared	0.3484		0.3272		0.2687		0.2611	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Appendix 6.1: Further descriptive statistics on persistence

Table A6.1.1: Transitions between mismatched, well-matched and not employed states over one-year intervals by year (%)

Status at t-1 / Status at *t*	Year (wave) which corresponds to *t*						
	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
A. Individuals employed at *t-1*							
Over-educated	*18.7*	*19.7*	*20.8*	*20.4*	*20.6*	*21.6*	*21.3*
Over-educated	61.1	62.6	63.2	65.5	66.4	63.6	65.9
Under-educated	2.7	2.5	2.9	2.4	2.2	3.4	3.7
Well-matched	7.1	6.7	7.4	8.9	8.0	9.2	8.2
Not employed	8.6	8.5	9.1	7.8	8.4	8.1	7.6
Not in-sample	20.5	19.8	17.5	15.5	14.9	15.8	14.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Under-educated	*9.2*	*10.6*	*10.1*	*10.8*	*10.8*	*10.9*	*10.9*
Over-educated	2.6	4.5	2.0	3.7	3.8	3.0	3.4
Under-educated	37.7	34.9	40.4	38.6	42.7	37.6	39.2
Well-matched	33.2	34.9	32.6	35.0	32.4	39.5	35.2
Not employed	6.2	7.2	6.2	8.1	6.3	7.1	6.3
Not in-sample	20.3	18.5	18.8	14.6	14.9	13.0	15.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Well-matched	*72.0*	*69.7*	*69.2*	*68.8*	*68.6*	*67.5*	*67.9*
Over-educated	1.8	2.1	1.9	1.8	2.2	2.4	2.4
Under-educated	3.9	4.3	4.1	4.3	3.9	4.4	4.2
Well-matched	71.6	75.2	74.5	77.2	77.7	76.4	77.5
Not employed	5.7	5.3	5.0	5.1	5.1	4.9	4.5
Not in-sample	17.0	13.2	14.6	11.5	11.1	11.8	11.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	6,218	5,870	5,938	5,817	6,120	6,258	6,279
B. Individuals not employed at *t-1*							
Not employed							
Over-educated	2.2	3.0	2.3	3.1	3.4	3.0	3.0
Under-educated	1.5	1.5	1.7	1.9	1.7	1.9	1.5
Well-matched	4.2	3.8	4.8	4.5	4.5	4.5	4.5
Not employed	72.9	75.0	75.1	76.1	77.8	77.2	79.0
Not in-sample	19.3	16.7	16.0	14.4	12.5	13.4	12.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	5,444	4,953	4,737	4,586	4,512	4,548	4,447

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on two-year panels of adjacent waves of the data; N refers to the number employed (in panel A) and not employed (in panel B) at *t-1*.
Italicised figures are the estimated incidences of educational mismatch at *t-1* (as reported in Chapter 4).
The remaining figures are proportions that sum to 100.0 for (sub-sections in) each column.
'Not in-sample' means that in period *t* the individuals either: turn 65 years of age; become a full-time student; become self-employed; or, attrit from the sample of respondents in the HILDA Survey data.

Table A6.1.2: Sources of over-education entries and exits over one-year intervals by year (%)

	Year (wave) which corresponds to *t*						
	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
A. Entering over-education: Individuals over-educated at *t*							
Well-matched at t-1							
Increased education	9.8	18.4	24.1	31.1	17.6	10.6	22.6
Changed occupation	86.6	78.2	76.0	67.6	80.2	85.6	75.5
Both	3.7	3.5	0.0	1.4	2.2	3.9	2.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	82	87	79	74	91	104	102
Under-educated at t-1							
Increased education	0.0	0.0	25.0	8.3	7.7	0.0	8.3
Changed occupation	73.3	92.6	75.0	70.8	84.6	95.0	91.7
Both	26.7	7.4	0.0	20.8	7.7	5.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	15	27	12	24	26	20	24
B. Exiting over-education: Individuals over-educated at *t-1*							
Well-matched at t							
Increased education	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Changed occupation	97.5	93.2	90.6	99.0	93.9	94.2	95.3
Both	2.5	6.8	9.4	1.0	6.1	5.8	4.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	80	74	85	100	98	120	106
Under-educated at t							
Increased education	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Changed occupation	96.8	96.3	97.0	92.6	92.6	97.7	100.0
Both	3.2	3.7	3.0	7.4	7.4	2.3	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	31	27	33	27	27	44	48

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on two-year panels of adjacent waves of the data; N refers to the number of individuals who enter over-education (in panel A) and exit over-education (in panel B) between *t-1* and *t*.
The remaining figures are proportions that sum to 100.0 for (sub-sections in) each column.

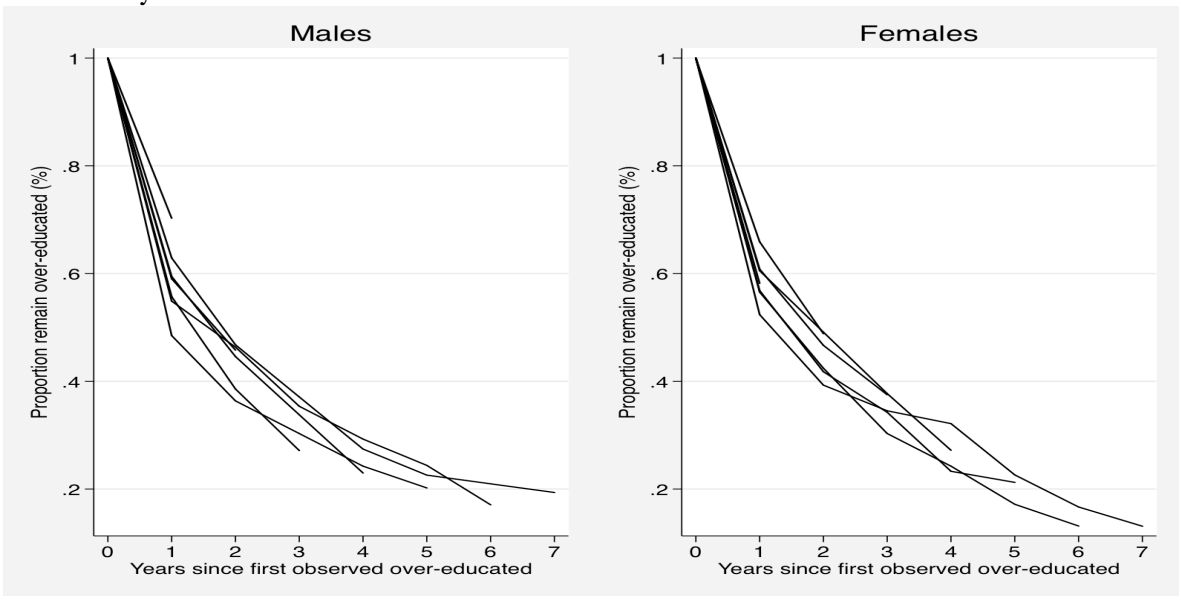
Table A6.1.3: Total number of periods observed over-educated by gender—Individuals in balanced panel sample (%)

	Males		Females	
	(I)	(II)	(I)	(II)
	In-sample at least once	In-sample each year	In-sample at least once	In-sample each year
8	3.9	8.5	3.6	9.7
7	1.7	1.4	1.9	1.5
6	2.3	2.1	3.0	2.0
5	2.2	2.0	3.5	0.9
4	2.1	1.1	3.5	3.0
3	4.4	1.6	4.9	2.9
2	4.5	2.4	5.5	1.7
1	8.8	5.6	9.6	3.6
0	70.1	75.4	64.5	74.8
Total	100.0	100.0	100.0	100.0
N	2,509	1,197	2,840	1,066

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that sum to 100.0 for each column and are weighted using longitudinal population weights (for the W1–W8 balanced panel) to make them representative of the Australian population.
Sample in (I) is individuals (in balanced panel sample) who are employed, between 15 and 64 years of age, not a full-time student and not self-employed at least once during the eight-year period (i.e., in-sample at least once); sample in (II) is individuals who satisfy these conditions in each of the eight years.

Figure A6.1.1: Survival functions—Proportion of individuals who remain over-educated by number of years since first observed over-educated



SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on the sample of individuals in the balanced panel.
Separate lines are presented for each year (wave) individuals' over-education spells were first observed: lines covering 7 years correspond to individuals whose spell was first observed in 2001 (W1), lines covering 6 years correspond to individuals whose spell was first observed in 2002 (W2), and so on.
New over-education spells are identified in 2001 (W1) using individuals' occupation tenure (i.e., new spells are defined for over-educated individuals with tenure less than one year), while for 2002 to 2007 (W2–W7) new spells are identified using individuals' over-education status in the previous year.
Survival functions treat exiting over-education as a terminal condition (i.e., can only occur once), which means they do not consider any individuals who later return to over-education; also, such exits occur because individuals either: become well-matched, under-educated or not employed; turn 65 years of age; become a full-time student; or, become self-employed.

Appendix 6.2: Wage penalty by duration over-educated

Table A6.2.1: Variation in wages and duration over-educated by gender

	Males					Females				
	Mean	Standard deviations			*N*	Mean	Standard deviations			*N*
		Overall	*Between*	*Within*			*Overall*	*Between*	*Within*	
ln(real hourly wage)	2.957	0.500	0.480	0.238	*21,459*	2.833	0.459	0.440	0.256	*21,850*
Duration over-educ.	1.149	4.072	3.307	2.012	*21,617*	1.268	3.872	3.143	1.982	*22,042*
Duration over-educated										
Less than 1 year	0.050	0.219	0.216	0.159	*21,617*	0.063	0.243	0.234	0.179	*22,042*
1–2 years	0.043	0.203	0.174	0.154	*21,617*	0.052	0.222	0.186	0.171	*22,042*
3–5 years	0.043	0.203	0.156	0.153	*21,617*	0.052	0.223	0.173	0.168	*22,042*
6–9 years	0.022	0.148	0.095	0.116	*21,617*	0.030	0.170	0.119	0.130	*22,042*
10+ years	0.040	0.196	0.150	0.114	*21,617*	0.046	0.209	0.161	0.124	*22,042*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Recall, *between* variation is variation across individuals, while the *within* is variation over time for each individual (Cameron and Trivedi, 2009).

Table A6.2.2: ORU earnings functions with duration over-educated—Fixed effects estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)		(II)		(I)		(II)	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
Duration over-educated	-0.008**	0.002			-0.004	0.003		
Duration over-educated-squared	0.001**	0.000			0.000	0.000		
Over-educated: Less than 1 year			-0.033	0.018			-0.078**	0.018
Over-educated: 1–2 years			-0.019	0.019			-0.059**	0.017
Over-educated: 3–5 years			-0.043*	0.018			-0.042*	0.018
Over-educated: 6–9 years			-0.051*	0.021			-0.031	0.020
Over-educated: 10+ years			-0.051*	0.023			-0.061**	0.021
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	-0.031	0.041	-0.030	0.041	-0.055	0.038	-0.052	0.037
Bachelor Degree	-0.038	0.034	-0.036	0.034	-0.069	0.036	-0.067	0.035
Advanced Diploma	-0.052	0.082	-0.053	0.082	-0.167*	0.075	-0.165*	0.073
Diploma	-0.073	0.054	-0.074	0.053	-0.122*	0.051	-0.112*	0.050
Certificate IV	0.034	0.062	0.034	0.061	-0.117*	0.051	-0.107*	0.051
Certificate III	0.004	0.059	0.000	0.058	-0.109*	0.053	-0.118*	0.052
Certificate II	-0.148	0.093	-0.150	0.093	-0.178*	0.073	-0.182*	0.073
Certificate I	-0.106	0.091	-0.113	0.091	0.062	0.140	0.026	0.146
Year 12	-0.082	0.061	-0.085	0.060	-0.128*	0.054	-0.131*	0.054
Year 11	-0.037	0.064	-0.045	0.064	-0.224**	0.078	-0.234**	0.077
Year 10 or below	-0.085	0.075	-0.092	0.075	-0.155*	0.064	-0.175**	0.064
Multiple qualifications	-0.005	0.018	-0.005	0.018	0.020	0.019	0.018	0.019
Experience	0.059**	0.019	0.058**	0.019	0.056**	0.016	0.054**	0.016
Experience-squared	-0.002**	0.000	-0.002**	0.000	-0.001**	0.000	-0.001**	0.000
Experience-cubed	0.000**	0.000	0.000**	0.000	0.000*	0.000	0.000*	0.000
Occupation tenure	0.003*	0.001	0.002	0.001	0.004**	0.001	0.003	0.001
Occupation tenure-squared	0.000*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Employer tenure	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
Employer tenure-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Years unemployed	-0.046	0.036	-0.046	0.036	-0.048	0.040	-0.045	0.040
English proficiency (base: English only)								
Good	-0.023	0.018	-0.022	0.018	0.011	0.023	0.010	0.023
Poor	-0.068	0.051	-0.068	0.051	-0.082	0.066	-0.077	0.067
L-T health condition	-0.005	0.007	-0.005	0.007	0.014	0.009	0.014	0.009
No. jobs last year	0.020**	0.006	0.020**	0.006	0.009	0.005	0.010*	0.005
% time employed last year	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% time unemployed last year	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000
Age (base: 15–19 years)								
20–24 years	0.106**	0.031	0.104**	0.031	0.164**	0.026	0.160**	0.026
25–29 years	0.091*	0.037	0.089*	0.037	0.173**	0.034	0.167**	0.034
30–34 years	0.080	0.042	0.077	0.042	0.173**	0.042	0.167**	0.042
35–39 years	0.076	0.045	0.073	0.045	0.140**	0.047	0.136**	0.047
40–44 years	0.107*	0.048	0.103*	0.048	0.120*	0.051	0.115*	0.051
45–49 years	0.115*	0.050	0.111*	0.050	0.105	0.055	0.100	0.054
50–54 years	0.119*	0.053	0.115*	0.053	0.107	0.059	0.104	0.059
55–59 years	0.107	0.057	0.102	0.057	0.083	0.064	0.080	0.064
60–64 years	0.158*	0.073	0.153*	0.073	0.076	0.072	0.075	0.072
Marital status (base: Legally married)								
Defacto	-0.023	0.014	-0.023	0.014	0.006	0.014	0.006	0.014
Separated	-0.009	0.020	-0.007	0.020	0.002	0.019	0.002	0.018
Divorced	-0.015	0.029	-0.015	0.029	0.011	0.023	0.013	0.023
Widowed	0.116	0.131	0.116	0.131	0.000	0.045	0.002	0.045
Single	-0.054**	0.019	-0.054**	0.019	-0.012	0.020	-0.012	0.020

Table A6.2.2 (continued)

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)		(II)		(I)		(II)	
Explanatory variables: Individual characteristics	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*
No. children had	-0.009	0.012	-0.009	0.012	-0.013	0.018	-0.011	0.018
Resident children by age								
No. aged 0–4 years	0.007	0.012	0.008	0.012	0.024	0.020	0.023	0.020
No. aged 5–14 years	-0.002	0.011	-0.002	0.011	-0.009	0.019	-0.010	0.020
No. aged 15–24 years	0.017	0.012	0.016	0.012	-0.017	0.020	-0.017	0.020
No. aged 25+ years	0.050*	0.023	0.048*	0.023	-0.034	0.026	-0.034	0.026
Non-resident children by age								
No. aged 0–4 years	0.014	0.023	0.013	0.023	0.025	0.051	0.025	0.049
No. aged 5–14 years	-0.005	0.015	-0.006	0.015	-0.011	0.032	-0.010	0.032
No. aged 15–24 years	0.009	0.012	0.008	0.012	-0.019	0.021	-0.019	0.021
No. aged 25+ years	0.031*	0.014	0.030*	0.014	-0.023	0.022	-0.023	0.022
Rest HH income	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rest HH income-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State/Territory of residence (base: NSW)								
Victoria	0.026	0.051	0.026	0.051	0.023	0.045	0.027	0.045
Queensland	-0.090**	0.033	-0.090**	0.033	0.039	0.042	0.040	0.042
South Australia	-0.033	0.055	-0.032	0.056	0.028	0.055	0.027	0.054
Western Australia	-0.008	0.050	-0.007	0.050	-0.009	0.076	-0.011	0.075
Tasmania	-0.160	0.097	-0.161	0.097	-0.072	0.082	-0.081	0.081
NT	0.080	0.072	0.081	0.073	0.071	0.102	0.070	0.101
ACT	0.026	0.038	0.024	0.038	0.007	0.065	0.009	0.063
Remoteness of area of residence (base: Major city)								
Inner regional	-0.024	0.023	-0.025	0.022	-0.062**	0.023	-0.062**	0.023
Outer regional	0.016	0.029	0.015	0.029	-0.084*	0.036	-0.084*	0.036
Remote (very remote)	0.038	0.054	0.035	0.054	-0.063	0.069	-0.061	0.068
No. homes lived in (base: Lived in 1 home)								
Lived in 2–3 homes	-0.010	0.026	-0.010	0.026	-0.026	0.027	-0.028	0.027
Lived in 4–9 homes	0.014	0.023	0.014	0.023	0.039*	0.019	0.040*	0.019
Lived in 10+ homes	0.005	0.039	0.007	0.039	-0.001	0.043	-0.003	0.043
Year controls (base: 2001)								
2002	0.012	0.018	0.012	0.018	-0.017	0.017	-0.017	0.017
2003	0.041	0.033	0.041	0.033	-0.020	0.029	-0.019	0.029
2004	0.065	0.049	0.066	0.049	-0.034	0.041	-0.032	0.041
2005	0.081	0.065	0.082	0.065	-0.035	0.055	-0.034	0.054
2006	0.090	0.081	0.092	0.081	-0.032	0.068	-0.030	0.068
2007	0.120	0.097	0.122	0.097	-0.023	0.081	-0.021	0.081
2008	0.117	0.112	0.119	0.112	-0.052	0.094	-0.050	0.094
Intercept term	2.467**	0.301	2.489**	0.301	2.235**	0.221	2.276**	0.220
N	20,636		20,636		21,035		21,035	
(No. individuals)	(5,323)		(5,323)		(5,449)		(5,449)	
Within R-squared	0.0675		0.0676		0.0407		0.0426	
Between R-squared	0.2036		0.2075		0.1512		0.1588	
Overall R-squared	0.1598		0.1636		0.1059		0.1132	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.
Robust standard errors reported; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Table A6.2.3: Over-education wage penalty by duration over-educated—ORU earnings functions, pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males				Females			
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
Duration over-educated	-0.023**		-0.026**		-0.028**		-0.025**	
	(0.003)		(0.004)		(0.002)		(0.004)	
Duration over-educ.-squared	0.001**		0.001**		0.001**		0.001**	
	(0.000)		(0.000)		(0.000)		(0.000)	
Over-educated: Less than 1 year		-0.136**		-0.171**		-0.156**		-0.129**
		(0.018)		(0.038)		(0.015)		(0.039)
Over-educated: 1–2 years		-0.165**		-0.211**		-0.145**		-0.162**
		(0.019)		(0.043)		(0.014)		(0.035)
Over-educated: 3–5 years		-0.195**		-0.178**		-0.152**		-0.178**
		(0.018)		(0.035)		(0.015)		(0.032)
Over-educated: 6–9 years		-0.170**		-0.208**		-0.140**		-0.157**
		(0.023)		(0.037)		(0.022)		(0.041)
Over-educated: 10+ years		-0.105**		-0.124**		-0.210**		-0.191**
		(0.024)		(0.034)		(0.019)		(0.031)
Balanced panel sample	No	No	Yes	Yes	No	No	Yes	Yes
N	20,288	20,288	6,688	6,688	20,806	20,806	6,200	6,200
(No. individuals)	(5,231)	(5,231)	(836)	(836)	(5,370)	(5,370)	(775)	(775)
R-squared	0.2995	0.3058	0.3044	0.3092	0.2875	0.2930	0.3286	0.3323
Wald test for joint significance of duration coefficients (H_0: All = 0) (p-value)		Reject H_0 (0.000)		Reject H_0 (0.000)		Reject H_0 (0.000)		Reject H_0 (0.000)
Wald test for equality of duration coefficients (H_0: All equal) (p-value)		Reject H_0 (0.007)		Do not reject H_0 (0.281)		Reject H_0 (0.017)		Do not reject H_0 (0.622)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *human capital measures*; *recent labour market experiences*; *demographic characteristics*; *family background*; and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories.

Complete results for these models are not presented, but are available from the author on request.

Additional Wald tests for equality of duration coefficients indicate the following statistically significant differences between coefficients: 'Less than 1 year' and '3–5 years' for males in (II); '3–5 years' and '10+ years' for males in (II); '6–9 years' and '10+ years' for males in (II); '3–5 years' and '10+ years' for females in (II); '6–9 years' and '10+ years' for females in (II); and, '6–9 years' and '10+ years' for males in (IV). Results from these Wald tests are not presented, but are available from the author on request.

Appendix 6.3: Details for dynamic panel probit model estimates

Table A6.3.1: Dynamic random effects probit model estimates for probability over-educated—Males

Dependent variable: Over-educated $_t$	Pooled probit		Heckman		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
Over-educated $_{t-1}$	2.848**	0.042	1.055**	0.096	1.400**	0.087	1.394**	0.088
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	0.113	0.394	-0.146	0.520	0.137	0.463	0.141	0.462
Bachelor Degree	0.398	0.357	0.197	0.459	0.456	0.424	0.480	0.420
Advanced Diploma	-1.888**	0.658	-0.538	1.229	-1.464**	0.754	-1.569**	0.768
Diploma	-1.528**	0.529	-1.956**	1.023	-1.129*	0.633	-1.123*	0.639
Certificate IV	-0.939	0.510	-1.957**	0.980	-0.805	0.614	-0.953*	0.619
Certificate III	-2.133**	0.502	-2.630**	0.975	-1.843**	0.613	-2.043**	0.619
Certificate II	-0.800	0.922	-2.337*	2.053	-0.680	1.143	-0.922	1.179
Year 12	-2.684**	0.490	-2.621**	1.008	-2.212**	0.616	-2.471**	0.619
Year 11	-1.638**	0.631	-2.392**	1.107	-1.434**	0.761	-1.661**	0.774
Experience	0.078	0.065	0.013	0.113	0.011	0.079	0.012	0.078
Experience-squared	0.001	0.003	0.000	0.005	0.002	0.004	0.002	0.004
Experience-cubed	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Occupation tenure	-0.020	0.011	-0.013	0.015	-0.015	0.013	-0.015	0.013
Occupation tenure-squared	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Years unemployed	0.228	0.391	0.523	0.688	0.238	0.451	0.211	0.451
L-T health condition	-0.064	0.080	-0.060	0.120	-0.045	0.097	-0.032	0.097
English proficiency (base: English only)								
Good	-0.032	0.192	0.031	0.301	-0.036	0.236	-0.031	0.236
Poor	0.122	0.600	0.031	1.048	0.078	0.775	0.087	0.781
Age (base: 15–19 years)								
20–24 years	-0.051	0.213	-0.096	0.485	-0.017	0.249	-0.125	0.246
25–29 years	-0.410	0.290	-0.300	0.602	-0.299	0.343	-0.411	0.339
30–34 years	-0.605	0.350	-0.430	0.688	-0.391	0.414	-0.512	0.411
35–39 years	-0.713	0.399	-0.503	0.754	-0.445	0.477	-0.575	0.474
40–44 years	-0.697	0.447	-0.535	0.814	-0.453	0.538	-0.580	0.534
45–49 years	-0.838	0.493	-0.544	0.868	-0.548	0.597	-0.676	0.593
50–54 years	-0.714	0.541	-0.487	0.925	-0.466	0.659	-0.604	0.655
55–59 years	-0.555	0.599	-0.314	1.007	-0.269	0.740	-0.412	0.737
60–64 years	-0.260	0.685	-0.125	1.135	0.055	0.856	-0.101	0.853
Ethnicity (base: Australian-born)								
ATSI	-0.122	0.186	0.269	0.503	-0.021	0.295	-0.114	0.298
Migrant ESB	-0.032	0.070	-0.072	0.153	-0.116	0.121	-0.024	0.119
Migrant Other	0.143	0.094	0.171	0.210	0.234*	0.161	0.167	0.161
Marital status (base: Married or defacto)								
Separated, divorced or widowed	0.068	0.177	0.015	0.247	0.061	0.213	0.064	0.212
Single	-0.003	0.126	-0.056	0.202	0.009	0.153	0.014	0.151
No. children had	0.113	0.078	0.060	0.117	0.112	0.096	0.116	0.095
Remoteness of area (base: Major city)								
Inner regional	0.063	0.155	0.035	0.233	0.034	0.180	0.043	0.180
Outer regional	-0.028	0.216	0.221	0.304	-0.021	0.246	0.004	0.246
Remote (very remote)	0.255	0.377	0.276	0.505	0.283	0.447	0.317	0.449
Year controls (base: 2002)								
2003	-0.003	0.073	-0.003	0.108	-0.013	0.090	-0.019	0.090
2004	-0.186*	0.079	-0.051	0.139	-0.171*	0.101	-0.186*	0.100

Table A6.3.1 (continued)

Dependent variable: Over-educated $_t$	Pooled probit		Heckman		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
2005	-0.215*	0.086	-0.110	0.180	-0.228**	0.114	-0.239**	0.114
2006	-0.168	0.093	0.015	0.224	-0.187*	0.129	-0.199*	0.129
2007	-0.228*	0.104	-0.054	0.274	-0.235*	0.147	-0.252*	0.147
2008	-0.263*	0.116	0.058	0.325	-0.273*	0.167	-0.282*	0.167
$\hat{e}_i$					0.715**	0.103		
Over-educated $_{initial}$							1.223**	0.179
Intercept term	-1.735**	0.203	-1.243**	0.531	-1.114**	0.315	-1.615**	0.324
Rho (ρ)			0.731**	0.042	0.471**	0.045	0.476**	0.045
Theta (θ)			1.220**	0.160				
N	13,002		16,314		12,904		13,002	
(No. individuals)	(3,183)		(3,226)		(3,160)		(3,183)	
Log likelihood	-2,317.22		-2,151.93		-2,219.14		-2,243.02	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels; significance tests for the Heckman, Orme and Wooldridge estimators are based on the original (rather than re-scaled) coefficients.

Re-scaled coefficients are the product of the original coefficients and estimate of $\sqrt{(1-\rho)}$ (using estimate of ρ derived from the model); original coefficients are not presented, but are available from the author on request.

All models also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).

See Table A6.3.3 in Appendix 6.3 for results of the initial conditions models (i.e., initial over-education status) that are estimated as part of the Heckman and Orme estimators.

Table A6.3.2: Dynamic random effects probit model estimates for probability over-educated—Females

Dependent variable: Over-educated $_t$	Pooled probit		Heckman		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
Over-educated $_{t-1}$	2.895**	0.043	1.323**	0.124	1.504**	0.092	1.441**	0.093
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	0.339	0.419	0.177	0.538	0.133	0.524	0.225	0.502
Bachelor Degree	0.416	0.407	-0.070	0.550	0.125	0.533	0.261	0.491
Advanced Diploma	-0.845	0.736	-1.280	1.245	-0.952	0.892	-0.837	0.916
Diploma	0.329	0.541	-0.260	0.817	0.054	0.702	0.185	0.686
Certificate IV	-0.067	0.522	-0.479	0.849	-0.363	0.683	-0.270	0.668
Certificate III	-2.180**	0.515	-2.797**	0.892	-2.416**	0.699	-2.452**	0.689
Certificate II	-1.676	0.898	-2.771**	1.373	-1.740*	1.102	-1.730*	1.113
Year 12	-1.561**	0.502	-1.848**	0.849	-1.746**	0.674	-1.731**	0.660
Year 11	-1.050	0.697	-1.237*	0.967	-1.023	0.851	-1.025	0.848
Experience	0.074	0.073	0.179*	0.135	0.018	0.090	0.022	0.091
Experience-squared	-0.005	0.004	-0.006	0.006	-0.003	0.005	-0.003	0.005
Experience-cubed	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Occupation tenure	0.002	0.012	0.002	0.018	0.001	0.014	0.006	0.015
Occupation tenure-squared	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001
Years unemployed	0.777*	0.332	0.834**	0.501	0.758**	0.381	0.705**	0.385
L-T health condition	-0.009	0.084	0.034	0.131	0.009	0.104	-0.001	0.105
English proficiency (base: English only)								
Good	-0.320	0.187	-0.214	0.303	-0.212	0.237	-0.259	0.240
Poor	-0.683	0.742	-0.261	1.314	-0.845	0.970	-0.590	0.968
Age (base: 15–19 years)								
20–24 years	0.371	0.242	-0.555	0.848	0.301	0.285	0.271	0.288
25–29 years	0.512	0.317	-0.596	0.924	0.499	0.380	0.376	0.381
30–34 years	0.509	0.380	-0.881	0.987	0.539	0.459	0.396	0.460
35–39 years	0.452	0.430	-1.037	1.039	0.483	0.520	0.363	0.521
40–44 years	0.446	0.472	-1.175	1.083	0.433	0.574	0.343	0.575
45–49 years	0.715	0.515	-1.002	1.126	0.703	0.629	0.620	0.631
50–54 years	0.713	0.562	-1.001	1.177	0.743	0.691	0.685	0.694
55–59 years	1.022	0.623	-0.670	1.243	1.099*	0.774	1.057	0.778
60–64 years	0.939	0.724	-0.816	1.384	0.916	0.912	0.885	0.920
Ethnicity (base: Australian-born)								
ATSI	0.174	0.165	0.140	0.453	0.200	0.269	0.125	0.276
Migrant ESB	0.111	0.071	0.050	0.154	0.141	0.120	0.141	0.122
Migrant Other	0.191*	0.091	0.500**	0.280	0.458**	0.161	0.242*	0.158
Marital status (base: Married or defacto)								
Separated, divorced or widowed	-0.075	0.186	-0.029	0.280	-0.093	0.230	-0.099	0.232
Single	0.280	0.145	0.144	0.247	0.173	0.175	0.189	0.174
No. children had	0.043	0.117	0.060	0.197	0.118	0.145	0.119	0.147
Remoteness of area (base: Major city)								
Inner regional	-0.093	0.187	0.128	0.306	-0.050	0.223	-0.036	0.225
Outer regional	0.159	0.271	0.033	0.500	0.112	0.347	0.095	0.348
Remote (very remote)	0.377	0.404	-0.326	0.826	0.136	0.514	0.122	0.520
Year controls (base: 2002)								
2003	0.078	0.076	0.052	0.111	0.043	0.094	0.048	0.094
2004	0.015	0.080	-0.085	0.145	-0.012	0.102	-0.004	0.103
2005	-0.066	0.086	-0.210	0.190	-0.104	0.114	-0.097	0.115
2006	-0.009	0.091	-0.303*	0.239	-0.065	0.126	-0.052	0.127
2007	-0.177	0.100	-0.353*	0.286	-0.220*	0.142	-0.211*	0.144
2008	-0.142	0.111	-0.346	0.334	-0.215	0.160	-0.204	0.162

Table A6.3.2 (continued)

Dependent variable: Over-educated $_t$	Pooled probit		Heckman		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
$\hat{e}_i$					0.704**	0.115		
Over-educated $_{initial}$							1.218**	0.201
Intercept term	-1.978**	0.209	-1.378**	0.538	-1.445**	0.330	-1.706**	0.338
Rho (ρ)			0.641**	0.062	0.475**	0.048	0.499**	0.047
Theta (θ)			1.424**	0.212				
N	11,670		14,931		11,608		11,670	
(No. individuals)	(3,077)		(3,121)		(3,057)		(3,077)	
Log likelihood	-2,171.61		-1,964.13		-2,086.01		-2,111.47	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels; significance tests for the Heckman, Orme and Wooldridge estimators are based on the original (rather than re-scaled) coefficients.

Re-scaled coefficients are the product of the original coefficients and estimate of $\sqrt{(1-\rho)}$ (using estimate of ρ derived from the model); original coefficients are not presented, but are available from the author on request.

All models also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).

See Table A6.3.3 in Appendix 6.3 for results of the initial conditions models (i.e., initial over-education status) that are estimated as part of the Heckman and Orme estimators.

Table A6.3.3: Probit models for probability over-educated in initial period—Initial conditions models results for Heckman and Orme estimators

Dependent variable: Over-educated $_{initial}$	Males				Females			
	Heckman		Orme		Heckman		Orme	
Explanatory variables: Individual characteristics $_{initial}$	*Coeff.*	*SE*	*Coeff.*	*SE*	*Coeff.*	*SE*	*Coeff.*	*SE*
Pre-sample information (Exclusion variables)								
No. jobs last year	0.215*	0.109	0.019	0.019	0.057	0.110	-0.016	0.018
% time employed last year	-0.015**	0.005	0.000	0.001	0.008*	0.004	-0.002**	0.001
% time unemployed last year	-0.003	0.009	0.002	0.001	0.012	0.007	0.000	0.001
Lived in 2–3 homes	0.094	0.191	0.068*	0.035	0.418*	0.198	0.083*	0.034
Lived in 4–9 homes	0.146	0.205	0.009	0.037	0.237	0.209	0.015	0.036
Lived in 10+ homes	-0.325	0.353	-0.037	0.061	-0.282	0.345	-0.228**	0.063
Father employed when aged 14	-0.479	0.496	-0.189*	0.076	-0.541	0.395	-0.156*	0.067
Father's employment unknown	-0.428	0.575	-0.192*	0.091	-0.814	0.535	0.026	0.085
Father unemployed for 6+ months	-0.089	0.251	0.136**	0.043	-0.072	0.240	0.050	0.042
Father's unemployment unknown	0.239	0.338	0.131**	0.049	0.437	0.390	0.156**	0.050
Mother employed when aged 14	0.016	0.141	-0.042	0.027	-0.016	0.143	-0.005	0.027
Mother's employment unknown	0.407	0.315	0.011	0.058	-0.776*	0.398	-0.214**	0.068
No. siblings	0.098**	0.035	0.053**	0.007	-0.015	0.041	0.023**	0.007
Eldest child	0.071	0.143	0.069**	0.026	-0.287	0.158	-0.023	0.027
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	-0.164	0.993	0.636**	0.205	0.767	1.077	0.018	0.246
Bachelor Degree	-0.365	0.988	0.697**	0.184	1.457	1.051	0.478*	0.227
Advanced Diploma	-0.497	2.582	-0.456	0.394	28.18	179.3	2.434**	0.481
Diploma	-2.194	2.467	0.739*	0.326	2.806	1.587	1.011**	0.300
Certificate IV	-2.697	2.813	-1.459**	0.442	0.179	1.626	0.605*	0.292
Certificate III	-5.765*	2.609	-2.047**	0.420	-2.604	1.584	-0.672*	0.289
Certificate II	-21.80	167.8	-19.09	497.5	-0.871	1.833	-0.711	0.416
Year 12	-4.829	2.559	-1.898**	0.413	0.217	1.428	0.015	0.262
Year 11	-4.821	2.875	-2.281**	0.469	-0.425	1.751	-0.485	0.335
Experience	-0.233	0.231	-0.059	0.041	0.201	0.252	0.054	0.042
Experience-squared	0.009	0.013	0.003	0.002	-0.008	0.015	-0.001	0.003
Experience-cubed	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Occupation tenure	-0.057	0.037	-0.029**	0.007	0.020	0.042	0.048**	0.008
Occupation tenure-squared	0.001	0.001	0.001**	0.000	-0.002	0.002	-0.002**	0.000
Years unemployed	-1.726*	0.813	-0.535**	0.157	-0.781	1.198	-0.881**	0.187
L-T health condition	0.367	0.280	0.195**	0.053	0.153	0.300	0.078	0.053
English proficiency (base: English only)								
Good	1.373*	0.551	0.361**	0.106	-0.423	0.544	-0.418**	0.102
Poor	1.235	1.136	0.305	0.244	0.600	1.619	1.314**	0.377
Age (base: 15–19 years)								
20–24 years	-0.396	0.707	-0.260*	0.105	-0.368	0.935	-0.068	0.111
25–29 years	-0.815	1.041	-0.502**	0.167	0.154	1.194	-0.071	0.169
30–34 years	-1.306	1.255	-0.741**	0.209	0.206	1.418	-0.243	0.213
35–39 years	-1.748	1.402	-1.144**	0.244	0.989	1.588	-0.001	0.248
40–44 years	-1.999	1.557	-1.211**	0.275	1.375	1.723	0.371	0.277
45–49 years	-2.307	1.713	-1.252**	0.306	1.589	1.885	0.390	0.307
50–54 years	-2.735	1.875	-1.917**	0.343	2.792	2.104	0.617	0.344
55–59 years	-2.336	2.129	-2.042**	0.393	2.658	2.338	0.565	0.395
60–64 years	-2.132	2.614	-2.247**	0.492	1.854	2.905	0.918	0.495
Ethnicity (base: Australian-born)								
ATSI	-0.054	0.770	0.054	0.112	-0.514	0.701	0.140	0.098
Migrant ESB	-0.464	0.247	-0.244**	0.045	-0.200	0.240	-0.021	0.044
Migrant Other	0.484	0.327	0.202**	0.057	1.188**	0.374	0.495**	0.053

Table A6.3.3 (continued)

Dependent variable: Over-educated $_{initial}$	Males				Females			
	Heckman		Orme		Heckman		Orme	
Explanatory variables: Individual characteristics $_{initial}$	*Coeff.*	*SE*	*Coeff.*	*SE*	*Coeff.*	*SE*	*Coeff.*	*SE*
Marital status (base: Married or defacto)								
Separated, divorced or widowed	0.217	0.446	-0.057	0.099	0.228	0.451	0.004	0.091
Single	0.228	0.384	-0.131	0.069	0.532	0.366	0.262**	0.066
No. children had	-0.043	0.217	0.009	0.043	0.716	0.380	0.445**	0.066
Remoteness of area (base: Major city)								
Inner regional	-0.161	0.483	-0.024	0.090	-0.252	0.477	-0.012	0.084
Outer regional	1.509*	0.643	0.455**	0.122	0.921	0.764	-0.032	0.121
Remote (very remote)	1.283	1.033	0.688**	0.207	-1.972	1.667	-0.445*	0.221
Intercept term	-1.269	1.051	-0.427**	0.155	-1.564	0.991	-0.538**	0.153
N	16,314		16,192		14,931		14,849	
(No. individuals)	(3,226)		(3,203)		(3,121)		(3,101)	
Log likelihood	-2,151.93		-6,846.22		-1,964.13		-6,936.48	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

All models also contain controls for *means of time-varying covariates* (where all variables except the pre-sample information and ethnicity exhibit variation over time).

Wald tests of the pre-sample information (or exclusion variables) coefficients indicate they are jointly statistically significant at the 5% level for the Heckman estimator and the 1% level for the Orme estimator.

Table A6.3.4: Dynamic random effects probit model estimates for probability over-educated—Orme and Wooldridge estimators using balanced panel sample

Dependent variable: Over-educated $_t$	Males				Females			
	Orme		Wooldridge		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
Over-educated $_{t-1}$	1.780**	0.141	1.694**	0.137	1.767**	0.150	1.688**	0.149
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	0.247	0.623	0.304	0.616	0.323	0.603	0.429	0.583
Bachelor Degree	0.546	0.550	0.582	0.541	-0.046	0.623	0.149	0.564
Advanced Diploma	-1.873	1.522	-1.786	1.566	-1.721	1.992	-2.215	2.670
Diploma	-2.449**	1.075	-2.304**	1.102	-0.424	0.914	-0.283	0.879
Certificate IV	-2.327**	1.011	-2.359**	1.021	-0.468	0.994	-0.258	0.996
Certificate III	-3.290**	1.004	-3.378**	1.012	-3.445**	1.045	-3.349**	1.044
Certificate II	-2.847*	1.791	-2.872	2.093	-2.928*	1.975	-2.769	2.065
Year 12	-3.399**	0.986	-3.549**	0.993	-2.136**	0.997	-2.016**	1.000
Year 11	-3.059**	1.181	-3.199**	1.222	-1.061	1.169	-0.884	1.162
Experience	-0.739	0.527	-0.714	0.534	0.254	0.578	0.249	0.590
Experience-squared	-0.002	0.006	-0.002	0.006	-0.010	0.008	-0.011*	0.008
Experience-cubed	0.000	0.000	0.000	0.000	0.000	0.000	0.000*	0.000
Occupation tenure	-0.026	0.017	-0.026*	0.017	0.003	0.020	0.007	0.020
Occupation tenure-squared	0.001	0.001	0.001	0.001	0.000	0.001	-0.001	0.001
Years unemployed	-0.409	1.019	-0.378	1.024	0.898	0.859	0.896	0.886
L-T health condition	-0.071	0.137	-0.055	0.138	0.079	0.165	0.044	0.166
English proficiency (base: English only)								
Good	-0.244	0.402	-0.262	0.412	-0.688*	0.436	-0.759*	0.447
Poor	-0.140	1.308	-0.079	1.412	-0.765	1.620	-0.638	1.810
Age (base: 15–19 years)								
20–24 years	-0.386	0.606	-0.748	0.553	-6.396	4753.0	-7.703	2451.0
25–29 years	-0.846	0.765	-1.225*	0.722	-6.497	4753.0	-7.992	2451.0
30–34 years	-1.150	0.872	-1.506*	0.832	-6.584	4753.0	-8.105	2451.0
35–39 years	-1.297	0.945	-1.652*	0.908	-6.825	4753.0	-8.313	2451.0
40–44 years	-1.289	1.005	-1.640*	0.970	-6.730	4753.0	-8.186	2451.0
45–49 years	-1.270	1.056	-1.621*	1.023	-6.368	4753.0	-7.810	2451.0
50–54 years	-1.019	1.111	-1.379	1.081	-6.126	4753.0	-7.549	2451.0
55–59 years	-0.683	1.203	-1.023	1.179	-5.570	4753.0	-6.983	2451.0
60–64 years	-0.432	1.407	-0.787	1.385	-5.569	4753.0	-6.996	2451.0
Ethnicity (base: Australian-born)								
ATSI	0.341	0.530	0.249	0.536	0.371	0.788	0.189	0.827
Migrant ESB	-0.140	0.179	-0.045	0.180	0.177	0.205	0.167	0.211
Migrant Other	0.279	0.263	0.198	0.268	0.284	0.300	0.080	0.308
Marital status (base: Married or defacto)								
Separated, divorced or widowed	0.083	0.282	0.093	0.285	0.053	0.342	0.058	0.345
Single	0.013	0.234	-0.011	0.235	0.330	0.330	0.324	0.321
No. children had	0.135	0.123	0.136	0.124	-0.091	0.246	-0.120	0.247
Remoteness of area (base: Major city)								
Inner regional	0.109	0.250	0.110	0.250	0.237	0.368	0.302	0.377
Outer regional	0.155	0.345	0.197	0.349	0.212	0.601	0.205	0.603
Remote (very remote)	0.091	0.558	0.157	0.583	-0.421	1.087	-0.373	1.069
Year controls (base: 2002)								
2003	0.849*	0.498	0.823*	0.507	-0.007	0.550	0.016	0.563
2004	1.506*	0.970	1.428	0.985	-0.173	1.075	-0.134	1.100
2005	2.215	1.463	2.110	1.486	-0.364	1.619	-0.316	1.658
2006	3.266*	1.941	3.126*	1.971	-0.391	2.149	-0.331	2.201
2007	3.941*	2.423	3.768*	2.460	-0.460	2.682	-0.401	2.749

Table A6.3.4 (continued)

Dependent variable: Over-educated $_t$	Males				Females			
	Orme		Wooldridge		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
2008	4.879*	2.906	4.671*	2.951	-0.481	3.208	-0.420	3.288
$\hat{e}_i$	0.745**	0.164			0.818**	0.213		
Over-educated $_{initial}$			1.348**	0.277			1.381**	0.360
Intercept term	-3.564*	2.086	-5.114**	2.144	-36.571	9086.6	-37.631	4833.5
Rho (ρ)	0.383**	0.079	0.405**	0.075	0.479**	0.082	0.506**	0.078
N	7,042		7,098		5,782		5,817	
(No. individuals)	(1,006)		(1,014)		(826)		(831)	
Log likelihood	-904.15		-911.01		-710.99		-727.93	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels; significance tests are based on the original (rather than re-scaled) coefficients.

Re-scaled coefficients are the product of the original coefficients and estimate of $\sqrt{(1-\rho)}$ (using estimate of ρ derived from the model); original coefficients are not presented, but are available from the author on request.

All models also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).

Results of initial conditions models for the Orme estimator are not presented, but are available from the author on request; Wald tests of the pre-sample information (or exclusion variables) coefficients indicate they are jointly statistically significant at the 1% level.

Table A6.3.5: Dynamic random effects probit model estimates for probability over-educated—Orme and Wooldridge estimators with extended specifications

Dependent variable: Over-educated $_t$	Males				Females			
	Orme		Wooldridge		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
Over-educated $_{t-1}$	1.403**	0.088	1.389**	0.089	1.530**	0.093	1.441**	0.094
Highest education (base: Postgraduate Degree)								
Graduate Diploma or Certificate	0.179	0.472	0.197	0.473	0.148	0.520	0.229	0.502
Bachelor Degree	0.529	0.447	0.576	0.446	0.065	0.535	0.192	0.501
Advanced Diploma	-1.415*	0.764	-1.485**	0.777	-1.084	0.905	-0.977	0.934
Diploma	-1.077*	0.643	-1.036*	0.649	-0.029	0.703	0.089	0.695
Certificate IV	-0.744	0.622	-0.878	0.627	-0.453	0.683	-0.370	0.675
Certificate III	-1.775**	0.630	-1.946**	0.637	-2.549**	0.713	-2.585**	0.713
Certificate II	-0.698	1.166	-0.953	1.202	-1.948*	1.130	-1.952*	1.147
Year 12	-2.133**	0.647	-2.364**	0.652	-1.917**	0.700	-1.893**	0.698
Year 11	-1.366*	0.777	-1.555**	0.789	-1.214	0.880	-1.216	0.886
Multiple qualifications	0.105	0.196	0.123	0.196	-0.130	0.201	-0.137	0.203
Experience	0.018	0.080	0.021	0.079	0.028	0.091	0.030	0.092
Experience-squared	0.002	0.004	0.002	0.004	-0.003	0.005	-0.004	0.005
Experience-cubed	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Occupation tenure	-0.009	0.014	-0.009	0.014	0.006	0.016	0.011	0.016
Occupation tenure-squared	0.000	0.000	0.000	0.000	0.000	0.001	-0.001	0.001
Employer tenure	-0.022	0.019	-0.021	0.019	-0.017	0.022	-0.016	0.022
Employer tenure-squared	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Years unemployed	0.176	0.455	0.163	0.457	0.764**	0.385	0.699*	0.390
L-T health condition	-0.044	0.098	-0.030	0.098	0.002	0.104	-0.010	0.105
English proficiency (base: English only)								
Good	-0.060	0.239	-0.053	0.239	-0.216	0.238	-0.263	0.241
Poor	0.028	0.785	0.030	0.792	-0.852	0.970	-0.586	0.965
Age (base: 15–19 years)								
20–24 years	-0.001	0.252	-0.108	0.248	0.298	0.286	0.268	0.290
25–29 years	-0.308	0.346	-0.420	0.342	0.489	0.381	0.368	0.384
30–34 years	-0.411	0.419	-0.533	0.415	0.533	0.460	0.391	0.463
35–39 years	-0.464	0.482	-0.598	0.478	0.483	0.521	0.362	0.525
40–44 years	-0.478	0.543	-0.612	0.540	0.434	0.574	0.343	0.578
45–49 years	-0.572	0.602	-0.707	0.599	0.709	0.629	0.627	0.634
50–54 years	-0.497	0.665	-0.647	0.662	0.760	0.692	0.700	0.698
55–59 years	-0.269	0.747	-0.425	0.744	1.101	0.775	1.058	0.782
60–64 years	0.075	0.863	-0.092	0.862	0.910	0.913	0.876	0.924
Ethnicity (base: Australian-born)								
ATSI	-0.011	0.299	-0.098	0.303	0.239	0.268	0.152	0.279
Migrant ESB	-0.108	0.123	-0.013	0.122	0.142	0.119	0.145	0.123
Migrant Other	0.263*	0.164	0.191	0.163	0.450**	0.160	0.240*	0.159
Marital status (base: Married or defacto)								
Separated, divorced or widowed	0.130	0.218	0.142	0.218	-0.047	0.240	-0.070	0.243
Single	0.047	0.155	0.052	0.153	0.180	0.179	0.187	0.178
No. children had	0.117	0.097	0.124	0.097	0.110	0.145	0.116	0.147
Rest HH income	0.000	0.000	0.000*	0.000	0.000	0.000	0.000	0.000
Rest HH income-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State/Territory of residence (base: NSW)								
Victoria	-0.019	0.353	-0.025	0.353	-0.064	0.471	0.016	0.472
Queensland	0.113	0.322	0.133	0.325	-0.272	0.379	-0.256	0.385
South Australia	-0.506	0.458	-0.476	0.457	-0.545	0.556	-0.506	0.558
Western Australia	0.569	0.474	0.568	0.476	-0.559	0.573	-0.477	0.581

Table A6.3.5 (continued)

Dependent variable: Over-educated $_t$	Males				Females			
	Orme		Wooldridge		Orme		Wooldridge	
Explanatory variables: Individual characteristics	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*	*Re-scaled coeff.*	*SE*
Tasmania	0.466	0.783	0.454	0.782	-0.525	0.717	-0.433	0.721
NT	0.008	0.614	-0.028	0.616	0.394	0.776	0.375	0.774
ACT	-0.781*	0.464	-0.806*	0.473	0.665	0.644	0.690	0.656
Remoteness of area (base: Major city)								
Inner regional	-0.010	0.185	-0.002	0.185	-0.046	0.234	-0.028	0.238
Outer regional	-0.049	0.259	-0.012	0.259	0.031	0.366	0.016	0.369
Remote (very remote)	0.125	0.482	0.195	0.488	-0.053	0.545	-0.046	0.554
Year controls (base: 2002)								
2003	-0.014	0.091	-0.018	0.090	0.048	0.094	0.050	0.095
2004	-0.170*	0.102	-0.184*	0.102	-0.007	0.102	-0.002	0.104
2005	-0.227**	0.116	-0.241**	0.116	-0.100	0.114	-0.096	0.116
2006	-0.190*	0.130	-0.205*	0.130	-0.057	0.126	-0.048	0.128
2007	-0.234*	0.148	-0.257*	0.149	-0.212*	0.142	-0.209*	0.145
2008	-0.268*	0.169	-0.283*	0.170	-0.206	0.160	-0.201	0.163
$\hat{e}_i$	0.725**	0.104			0.693**	0.116		
Over-educated $_{initial}$			1.240**	0.182			1.220**	0.204
Intercept term	-1.056**	0.347	-1.653**	0.358	-1.319**	0.363	-1.631**	0.376
Rho (ρ)	0.475**	0.045	0.482**	0.044	0.465**	0.049	0.497**	0.047
N	12,894		12,992		11,604		11,666	
(No. individuals)	(3,158)		(3,181)		(3,057)		(3,077)	
Log likelihood	-2,202.65		-2,225.95		-2,075.12		-2,098.51	

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels; significance tests are based on the original (rather than re-scaled) coefficients.

Re-scaled coefficients are the product of the original coefficients and estimate of $\sqrt{(1-\rho)}$ (using estimate of ρ derived from the model); original coefficients are not presented, but are available from the author on request.

All models also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except ethnicity exhibit variation over time).

Results of initial conditions models for the Orme estimator are not presented, but are available from the author on request; Wald tests of the pre-sample information (or exclusion variables) coefficients indicate they are jointly statistically significant at the 1% level.

Table A6.3.6: Estimates of state dependence in over-education—Robustness analyses

	Males			Females		
	(I) Raw data	(II) Pooled probit	(III) Dynamic RE probit (Wooldridge)	(I) Raw data	(II) Pooled probit	(III) Dynamic RE probit (Wooldridge)
A. Balanced panel sample						
$Pr(O_t=1 \mid O_{t-1}=0)$ [P_0]	0.019	0.024	0.045	0.018	0.024	0.055
$Pr(O_t=1 \mid O_{t-1}=1)$ [P_1]	0.879	0.827	0.394	0.900	0.854	0.403
State dependence [$P_1 - P_0$]	0.860	0.804	0.349	0.882	0.831	0.348
Ratio [P_1 / P_0]	46.3	34.5	8.8	50.0	35.6	7.3
N	8,379	7,098	7,098	7,462	5,817	5,817
(No. individuals)	(1,197)	(1,014)	(1,014)	(1,066)	(831)	(831)
B. Extended specifications						
$Pr(O_t=1 \mid O_{t-1}=0)$ [P_0]	0.027	0.033	0.062	0.028	0.037	0.081
$Pr(O_t=1 \mid O_{t-1}=1)$ [P_1]	0.845	0.801	0.366	0.865	0.815	0.425
State dependence [$P_1 - P_0$]	0.818	0.767	0.305	0.837	0.779	0.344
Ratio [P_1 / P_0]	31.3	24.3	5.9	30.9	22.0	5.2
N	17,359	12,992	12,992	16,637	11,666	11,666
(No. individuals)	(4,260)	(3,181)	(3,181)	(4,240)	(3,077)	(3,077)

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Figures in column (I) are based on the raw data, while figures in columns (II) and (III) are derived from results of the dynamic models for over-education.

Estimates differ across the columns for the following reasons: estimates in (I) do not account for any individual heterogeneity; estimates in (II) control for observed individual heterogeneity; and, estimates in (III) control for observed and unobserved individual heterogeneity (and the endogeneity of each individual's initial over-education status). Hence, column (III) presents the most reliable estimates of state dependence in over-education.

Appendix 6.4: Prior over-education and wages of well-matched individuals

Table A6.4.1: Wage effects of prior over-education among well-matched individuals—ORU earnings functions, pooled OLS estimates

Dependent variable: ln(real hourly wage)	Males		Females	
	(I)	(III)	(I)	(III)
A. Sample: Individuals well-matched at *t* (two-year panels)				
Over-educated $_{t-1}$	-0.087**	-0.108**	-0.075**	-0.082**
	(0.025)	(0.031)	(0.020)	(0.024)
Restricted sample: Individuals who changed occupation between *t-1* and *t*	No	Yes	No	Yes
N	11,301	1,112	10,882	1,112
(No. individuals)	(3,041)	(866)	(3,012)	(864)
R-squared	0.2830	0.4377	0.3189	0.4214
B. Sample: Individuals well-matched at *t* and *t-1* (three-year panels)				
Over-educated $_{t-2}$	-0.130**	-0.153**	-0.039	-0.070*
	(0.026)	(0.031)	(0.023)	(0.028)
Restricted sample: Individuals who changed occupation between *t-2* and *t-1*	No	Yes	No	Yes
N	7,942	1,042	7,551	995
(No. individuals)	(2,248)	(654)	(2,238)	(628)
R-squared	0.2850	0.4890	0.3198	0.4724

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Results in panel A are based on analysis of pooled two-year panels of adjacent waves of the data, while results in panel B are based on analysis of pooled three-year panels of adjacent waves of the data.

** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *human capital measures*; *recent labour market experiences*; *demographic characteristics*; *family background*; and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories.

Table A6.4.2: Wage effects of prior over-education among well-matched individuals—ORU earnings functions, fixed effects estimates

Dependent variable: ln(real hourly wage)	Males		Females	
	(I)	(II)	(I)	(II)
Over-educated $_{t-1}$	-0.002	-0.016	-0.035	-0.045*
	(0.019)	(0.018)	(0.020)	(0.020)
Omit variables with little within variation	No	Yes	No	Yes
N	11,467	11,748	10,976	11,187
(No. individuals)	(3,094)	(3,198)	(3,046)	(3,134)
Within R-squared	0.0709	0.0512	0.0331	0.0276
Between R-squared	0.1245	0.0605	0.0498	0.0114
Overall R-squared	0.0961	0.0536	0.0341	0.0134

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

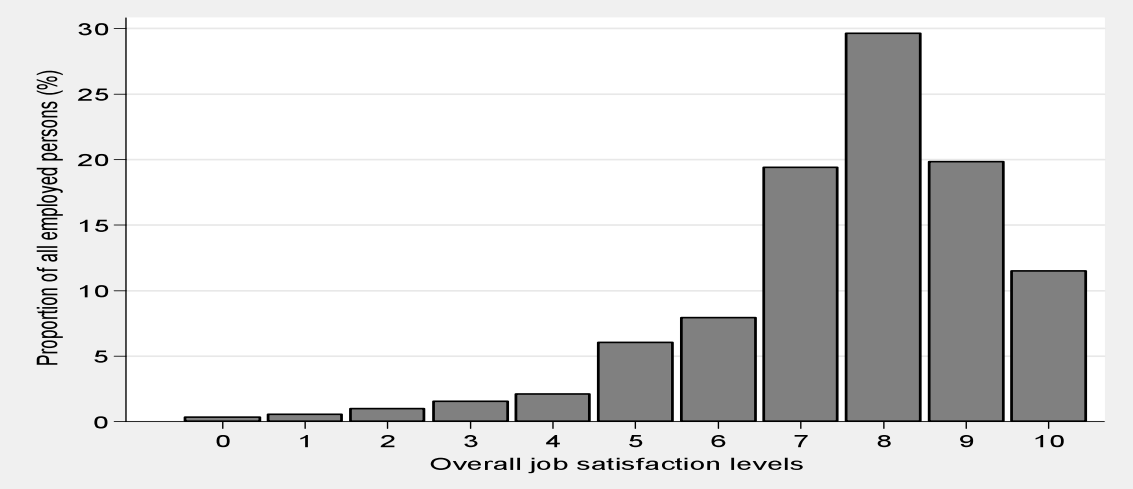
NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

Models in (I) contain the following categories of controls: *human capital measures; recent labour market experiences*; *demographic characteristics* (excluding ethnicity and number of homes lived in during past 10 years); and, *year dummies*. See Notes to Table 5.3 for exact list of variables contained in these categories. The variables with little within variation omitted from models in (II) are: highest education; experience; occupation tenure; employer tenure; English proficiency; State/Territory of residence; and, remoteness of area of residence.

Appendix 7.1: Distributions of overall job satisfaction levels and intentions to quit

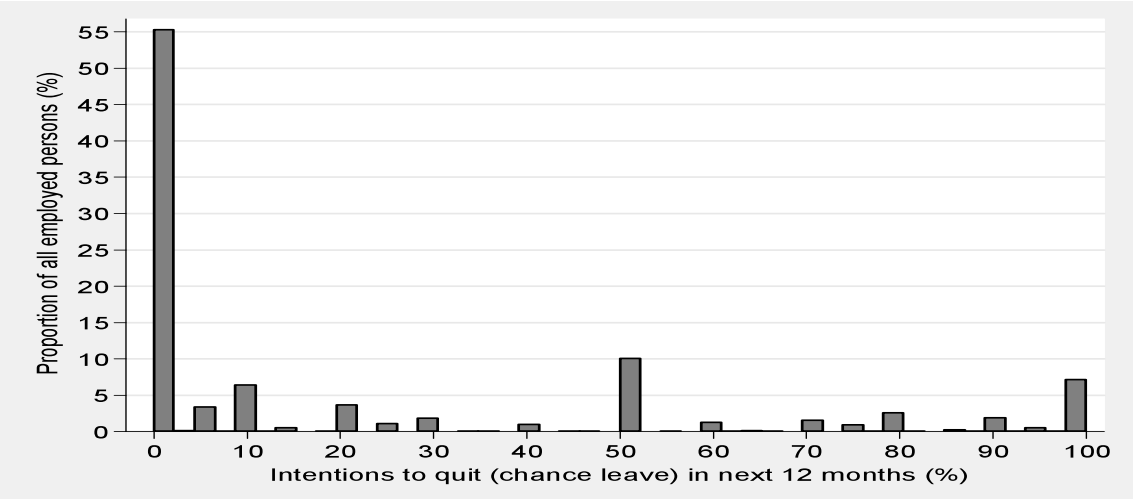
Figure A7.1.1: Distribution of overall job satisfaction levels—Employed individuals aged 15–64 years, excluding full-time students and self-employed



SOURCE: Author's calculations using HILDA Survey data (Release 8.0); specifically, the variables for overall job satisfaction level ('_jbmsall') in each year.

NOTES: Sample restricted to individuals who are between 15 and 64 years of age, not a full-time student and not self-employed; data pooled for the eight-year period from 2001 to 2008.

Figure A7.1.2: Distribution of intentions to quit in next 12 months—Employed individuals aged 15–64 years, excluding full-time students and self-employed



SOURCE: Author's calculations using HILDA Survey data (Release 8.0); specifically, the variables for intentions to quit (or chance of voluntarily leaving) in next 12 months ('_jbmplej') in each year.

NOTES: Sample restricted to individuals who are between 15 and 64 years of age, not a full-time student and not self-employed; data pooled for the eight-year period from 2001 to 2008.

Appendix 7.2: Estimates of voluntary over-education

Table A7.2.1: Incidence of voluntary over-education by gender and year using various thresholds for 'highly satisfied' and 'highly unlikely to quit'—Employed individuals aged 15–64 years, excluding full-time students and self-employed (%)

'Highly satisfied'	'Highly unlikely to quit'	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Males									
Job satisfaction=9–10	Chance quit=0%	2.4	3.2	3.5	2.6	3.0	2.5	3.1	2.9
Job satisfaction=9–10	Chance quit=0–5%	2.7	3.3	3.7	3.0	3.0	2.7	3.3	3.2
Job satisfaction=9–10	Chance quit=0–10%	2.9	3.5	3.9	3.2	3.3	2.9	3.5	3.5
Job satisfaction=8–10	Chance quit=0%	4.4	5.5	6.0	5.1	5.3	5.0	5.7	5.7
Job satisfaction=8–10	Chance quit=0–5%	4.8	6.0	6.3	5.9	5.6	5.7	6.1	6.5
Job satisfaction=8–10	Chance quit=0–10%	5.3	6.6	7.0	6.6	6.6	6.2	6.7	7.6
Job satisfaction=7–10	Chance quit=0%	6.2	7.0	7.4	6.6	6.4	6.1	7.4	6.9
Job satisfaction=7–10	Chance quit=0–5%	6.8	7.7	7.9	7.5	6.8	7.0	8.0	7.9
Job satisfaction=7–10	Chance quit=0–10%	7.6	8.7	9.1	8.5	8.1	7.9	9.1	9.5
Females									
Job satisfaction=9–10	Chance quit=0%	4.8	3.9	4.7	3.9	4.2	4.6	5.0	5.1
Job satisfaction=9–10	Chance quit=0–5%	5.1	4.1	4.8	4.0	4.5	4.8	5.3	5.2
Job satisfaction=9–10	Chance quit=0–10%	5.5	4.5	5.1	4.2	4.6	5.1	5.7	5.6
Job satisfaction=8–10	Chance quit=0%	7.7	6.6	7.9	7.3	8.1	8.2	9.1	9.3
Job satisfaction=8–10	Chance quit=0–5%	8.2	7.2	8.5	7.8	8.7	8.7	9.9	9.8
Job satisfaction=8–10	Chance quit=0–10%	8.9	7.9	9.2	8.7	9.5	9.6	10.8	11.2
Job satisfaction=7–10	Chance quit=0%	9.6	8.2	10.1	9.0	9.7	10.2	10.5	11.1
Job satisfaction=7–10	Chance quit=0–5%	10.1	8.9	10.8	9.9	10.5	10.9	11.5	11.8
Job satisfaction=7–10	Chance quit=0–10%	10.9	10.0	11.7	11.0	11.5	12.1	12.9	13.3

SOURCE: Author's calculations based on HILDA Survey data (Release 8.0).

NOTES: Figures are proportions that are weighted using cross-sectional population weights to make them representative of the Australian population of employed individuals who are between 15 and 64 years of age, not a full-time student and not self-employed.

Appendix 7.3: Definition of variables and descriptive statistics

Table A7.3.1: Definition of dependent variables (job attributes) used in analyses

Variable name	Description
ln(real hourly wage)	Natural logarithm of individual's real hourly wage in (main) job (for details on derivation see Table A5.5.1 in Appendix 5.1)
Probability of job loss	Employment prospects over next 12 months: Chance of losing (main) job (i.e., get retrenched, fired or not have contract renewed) (measured as probabilities from 0% (no chance) to 100% (absolute certainty))
Hours (per week)	Hours per week usually worked in (main) job
Travel time (hours per week)	Number of hours spent travelling to and from work in a typical week (derived from HILDA Self-Completion Questionnaire (SCQ))
Satisfaction with pay	Level of satisfaction with aspects of (main) job: Total pay (measured on 0 (totally dissatisfied) to 10 (totally satisfied) scale)
Satisfaction with job security	Level of satisfaction with aspects of (main) job: The job security (measured 0 to 10, as above)
Satisfaction with hours	Level of satisfaction with aspects of (main) job: The hours worked (measured 0 to 10, as above)
Satisfaction with work-life balance	Level of satisfaction with aspects of (main) job: The flexibility available to balance work and non-work commitments (measured 0 to 10, as above)
Satisfaction with the work	Level of satisfaction with aspects of (main) job: The work itself (measured 0 to 10, as above)
Flexibility	Aspects of (main) job: Flexibility – Have a lot of freedom to decide *when* to do work (derived from HILDA SCQ) (measured on 1 (strongly disagree) to 7 (strongly agree) scale)
Autonomy	Aspects of (main) job: Autonomy – Have a lot of freedom to decide *how* do work (derived from HILDA SCQ) (measured 1 to 7, as above)
Input	Aspects of (main) job: Input – Have a lot of say about what happens at work (derived from HILDA SCQ) (measured 1 to 7, as above)
Stressfulness	Aspects of (main) job: Stressfulness – Job is more stressful than had ever imagined (derived from HILDA SCQ) (measured 1 to 7, as above)
Complexity	Aspects of (main) job: Complexity – Job is complex and difficult (derived from HILDA SCQ) (measured 1 to 7, as above)
Learning of new skills	Aspects of (main) job: Learning – Job often requires learning of new skills (derived from HILDA SCQ) (measured 1 to 7, as above)
Use of existing skills	Aspects of (main) job: Skill use – Job *uses* many of my skills and abilities (derived from HILDA SCQ) (measured 1 to 7, as above)
Permanent contract	Individual employed on a permanent or ongoing basis in (main) job
Preferred number of hours	Individual works their preferred number of hours (i.e., if could choose hours worked each week, and taking into account how this would affect income, then would choose about the same hours as currently worked)
Flexible start/finish times	Individual has access to flexible start and finish times in (main) job (derived from HILDA SCQ)

SOURCE: Summerfield (2010) and the HILDA Online Data Dictionary (see http://www.melbourneinstitute.com/hildaddictionary/onlinedd/default.aspx).

NOTES: Due to non-completion of the HILDA Self-Completion Questionnaire (SCQ) among individuals who responded to the Person Questionnaire, the variables derived from the SCQ are subject to a higher degree of missing information.

Table A7.3.2: Descriptive statistics for dependent variables by over-education status

Variable	Well-matched			Voluntarily over-educated			Involuntarily over-educated		
	mean	s.d	*N*	mean	s.d	*N*	mean	s.d	*N*
Real hourly wage	21.08	14.57	*33,953*	18.13	10.96	*3,472*	16.89	15.51	*6,176*
ln(real hourly wage)	2.937	0.482	*33,742*	2.792	0.472	*3,457*	2.721	0.453	*6,110*
Probability of job loss	9.601	20.41	*33,923*	5.837	16.51	*3,472*	14.42	23.78	*6,151*
Hours (per week)	38.29	13.35	*33,960*	33.99	13.64	*3,474*	32.55	13.38	*6,177*
Travel time (hours per week)	3.915	3.890	*30,675*	3.682	4.046	*3,177*	3.725	3.995	*5,493*
Satisfaction with pay	6.993	2.108	*33,966*	7.594	1.802	*3,474*	6.173	2.242	*6,179*
Satisfaction with job security	8.051	2.109	*33,971*	8.499	1.721	*3,474*	7.323	2.374	*6,176*
Satisfaction with hours	7.225	2.117	*33,981*	8.106	1.672	*3,474*	6.545	2.268	*6,181*
Satisfaction with work-life balance	7.318	2.390	*33,966*	8.333	1.784	*3,474*	6.858	2.467	*6,177*
Satisfaction with the work	7.659	1.841	*33,986*	8.398	1.338	*3,475*	6.497	2.150	*6,181*
Flexibility	3.354	1.890	*30,458*	3.442	1.941	*3,151*	3.022	1.804	*5,369*
Autonomy	4.723	1.690	*30,466*	4.694	1.747	*3,153*	4.049	1.802	*5,367*
Input	4.284	1.715	*30,451*	4.269	1.759	*3,147*	3.533	1.738	*5,362*
Stressfulness	3.291	1.669	*30,464*	2.697	1.556	*3,151*	3.204	1.687	*5,369*
Complexity	4.164	1.840	*30,455*	3.427	1.857	*3,146*	3.277	1.754	*5,364*
Learning of new skills	4.718	1.741	*30,461*	4.465	1.872	*3,146*	4.032	1.841	*5,364*
Use of existing skills	5.495	1.423	*30,452*	5.346	1.578	*3,146*	4.504	1.825	*5,370*
Permanent contract	0.748	0.434	*33,977*	0.677	0.468	*3,475*	0.563	0.496	*6,186*
Preferred number of hours	0.572	0.495	*33,957*	0.664	0.473	*3,473*	0.513	0.500	*6,176*
Flexible start/finish times	0.522	0.500	*28,001*	0.587	0.492	*2,864*	0.524	0.499	*4,820*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Voluntary over-education identified using the (preferred) definition outlined in Section 7.3 (i.e., over-educated individuals with a job satisfaction level of 8 or above and a chance of quitting of 5 per cent or below are deemed voluntarily over-educated).

's.d' refers to standard deviations and '*N*' column reports total number of individuals with non-missing values for the variable.

Appendix 7.4: Robustness analyses for differences in job attributes

Table A7.4.1: Variation in voluntary and involuntary over-education identifiers

	Mean	Standard deviations			*N*
		Overall	*Between*	*Within*	
All individuals					
Voluntarily over-educated	0.080	0.271	0.223	0.175	*43,659*
Involuntarily over-educated	0.142	0.349	0.323	0.210	*43,659*
Males					
Voluntarily over-educated	0.066	0.248	0.209	0.159	*21,617*
Involuntarily over-educated	0.133	0.339	0.315	0.199	*21,617*
Females					
Voluntarily over-educated	0.093	0.290	0.234	0.190	*22,042*
Involuntarily over-educated	0.150	0.357	0.331	0.220	*22,042*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Recall, *between* variation is variation across individuals, while the *within* is variation over time for each individual (Cameron and Trivedi, 2009).

Table A7.4.2: Estimated differences in job attributes between over-educated and well-matched individuals—Sensitivity analyses using alternative estimators

	(I) Over-educated		(II) Voluntarily over-educated		Involuntarily over-educated		*N*
Dependent variables	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	*Coeff.*	*Robust SE*	
A. Continuous-valued job attributes: *Treated as ordinal-valued* (Fixed effects ordered logit estimates)							
Probability of job loss [a]	0.022	(0.095)	-0.515**	(0.156)	0.087	(0.096)	*82,576*
B. Ordinal-valued job attributes: *Alternative estimator* (Random effects ordered probit estimates)							
Satisfaction with pay [b]	-0.111**	(0.021)	0.641**	(0.035)	-0.273**	(0.022)	*41,993*
Satisfaction with job security [b]	-0.090**	(0.022)	0.554**	(0.037)	-0.219**	(0.023)	*41,994*
Satisfaction with hours [b]	-0.032	(0.021)	0.817**	(0.035)	-0.208**	(0.022)	*42,009*
Satisfaction with work-life balance [b]	0.057**	(0.022)	0.858**	(0.037)	-0.103**	(0.022)	*41,991*
Satisfaction with the work [b]	-0.260**	(0.021)	0.786**	(0.036)	-0.480**	(0.022)	*42,014*
Flexibility [c]	-0.136**	(0.024)	0.093*	(0.038)	-0.189**	(0.025)	*37,741*
Autonomy [c]	-0.268**	(0.023)	0.095*	(0.037)	-0.349**	(0.024)	*37,749*
Input [c]	-0.293**	(0.023)	0.071	(0.037)	-0.376**	(0.024)	*37,724*
Stressfulness [c]	-0.285**	(0.023)	-0.692**	(0.038)	-0.197**	(0.023)	*37,747*
Complexity [c]	-0.635**	(0.024)	-0.600**	(0.038)	-0.642**	(0.025)	*37,729*
Learning of new skills [c]	-0.507**	(0.024)	-0.268**	(0.038)	-0.562**	(0.025)	*37,734*
Use of existing skills [c]	-0.534**	(0.024)	-0.154**	(0.039)	-0.618**	(0.025)	*37,729*
C. Ordinal-valued job attributes: *Treated as continuous-valued* (Fixed effects estimates)							
Satisfaction with pay [b]	-0.207**	(0.063)	0.723**	(0.077)	-0.388**	(0.063)	*41,993*
Satisfaction with job security [b]	-0.172**	(0.061)	0.474**	(0.076)	-0.298**	(0.062)	*41,994*
Satisfaction with hours [b]	-0.068	(0.061)	0.934**	(0.076)	-0.262**	(0.061)	*42,009*
Satisfaction with work-life balance [b]	0.114	(0.068)	1.041**	(0.083)	-0.066	(0.069)	*41,991*
Satisfaction with the work [b]	-0.423**	(0.055)	0.615**	(0.065)	-0.624**	(0.055)	*42,014*
Flexibility [c]	-0.118*	(0.052)	0.111	(0.070)	-0.165**	(0.053)	*37,741*
Autonomy [c]	-0.256**	(0.049)	0.079	(0.065)	-0.324**	(0.050)	*37,749*
Input [c]	-0.300**	(0.049)	0.049	(0.065)	-0.371**	(0.049)	*37,724*
Stressfulness [c]	-0.315**	(0.047)	-0.704**	(0.063)	-0.236**	(0.048)	*37,747*
Complexity [c]	-0.551**	(0.048)	-0.484**	(0.067)	-0.564**	(0.048)	*37,729*
Learning of new skills [c]	-0.457**	(0.051)	-0.239**	(0.067)	-0.501**	(0.052)	*37,734*
Use of existing skills [c]	-0.573**	(0.045)	-0.311**	(0.058)	-0.626**	(0.046)	*37,729*
D. Binary-valued job attributes: *Alternative estimator* (Conditional fixed effects logit estimates)							
Permanent contract	-0.521**	(0.078)	-0.259*	(0.121)	-0.560**	(0.079)	*17,445*
Preferred number of hours	-0.144*	(0.064)	0.285**	(0.097)	-0.222**	(0.065)	*28,129*
Flexible start/finish times	-0.044	(0.084)	0.379**	(0.123)	-0.136	(0.086)	*16,260*

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: ** and * indicate statistical significance at the 1% and 5% levels.

Robust standard errors reported in parentheses; these cluster on individual identifiers to account for potential serial correlation in the errors of individuals with multiple observations.

All models contain the following categories of controls: *gender*; *human capital measures*; *recent labour market experiences*; *demographic characteristics*; and, *year dummies*. See Notes to Table 7.2 for exact list of variables contained in these categories. Controls for gender and ethnicity drop out of models in panels A, C and D, and models in panel B also contain controls for *means of time-varying covariates* (the Mundlak-Chamberlain controls) (where all variables except gender and ethnicity exhibit variation over time).

(a) Probability of job loss converted from being measured on 0-100 scale to 0-10 scale, where 0 represents a 0 per cent chance, 1 is 1-10 per cent chance, 2 is 11-20 per cent chance, …, and 10 is 91-100 per cent chance.

(b) Job satisfaction levels measured on scale from 0 (totally dissatisfied) to 10 (totally satisfied).

(c) Job attributes measured on scale from 1 (strongly disagree) to 7 (strongly agree).

Appendix 7.5: Descriptive statistics on persistence and voluntary over-education

Table A7.5.1: Persistent over-education (using duration over-educated) by voluntariness (at time *t*) and gender (%)

	2001 (W1)	2002 (W2)	2003 (W3)	2004 (W4)	2005 (W5)	2006 (W6)	2007 (W7)	2008 (W8)
Males								
Over-educated at t	*16.2*	*18.3*	*19.3*	*18.5*	*18.2*	*19.4*	*18.7*	*19.6*
% persistently over-educated	29.6	30.5	28.7	31.4	31.1	35.1	35.6	33.1
% voluntarily over-educated at *t*	35.1	47.8	42.0	37.3	33.7	36.1	40.6	39.6
N	490	519	540	510	517	567	563	596
Females								
Over-educated at t	*21.5*	*21.3*	*22.5*	*22.4*	*23.4*	*24.1*	*24.1*	*23.8*
% persistently over-educated	31.4	31.9	31.4	30.1	31.1	31.7	34.2	31.9
% voluntarily over-educated at *t*	46.0	43.8	53.2	40.7	44.4	42.1	49.3	53.2
N	640	581	616	619	704	742	728	732

SOURCE: Author's calculations using HILDA Survey data (Release 8.0).

NOTES: Analyses based on the restricted sample (as defined in Chapter 3) in each year (i.e., unbalanced panels are used); N refers to the number of individuals over-educated at *t*.

Italicised figures are the estimated incidences of over-education reported in Chapter 4. Individuals considered *persistently over-educated* if they have current duration over-educated of 6 or more years (where duration over-educated is approximated using individuals' occupation tenure); proportions above correspond to the figures reported in Table 6.3 in Chapter 6. Individuals identified as *voluntarily over-educated* using the preferred definition of voluntary over-education, as presented in panel A of Table 7.1 in Chapter 7.