

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
bibby, Yan

Title:
Plausible values: How many for plausible results?

Date:
2019

Persistent Link:
<https://hdl.handle.net/11343/235798>

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.

Plausible Values: How Many for Plausible Results?

Yan Bibby

orcid.org/0000-0002-5346-8964

Dissertation submitted in total fulfilment of the
requirements for the degree of Doctor of Philosophy

March 2020

Melbourne Graduate School of Education

University of Melbourne

Abstract

A common purpose of large-scale assessments is to describe the performance of populations of students, with the aim to refine education policies and strategies and improve education systems. Large-scale assessments are often administered to a sample of students to reduce costs. Large scale assessment may also utilise rotated test designs to reduce the burden of a long test on individual students.

The plausible value imputation technique was developed in the 1980's to analyse NAEP 1983-84 results to overcome the issues of using individual student estimates to make inferences regarding population parameters. Since then the plausible value technique has been widely used in large-scale assessments as a computational approach for estimating population parameters. When first implemented, five imputations were considered sufficient for all settings. In recent years, however, many researchers have recommended increases to well beyond five imputations. In this dissertation, two simulation studies and two empirical data studies were conducted in order to investigate the relationship between the number of imputations and the properties of population parameter estimates.

The results from the simulation studies show that there is a small bias in population parameter estimates, increasing the number of plausible values used from 1 to 20 did not reduce the magnitude of the bias for the simulated conditions. The RMSEs of population parameter estimates showed a small decreasing trend when an increased number of plausible values were used. A similar decreasing trend was also observed in the estimated standard errors of population parameter estimates with an increase in the number of plausible values used. It was found that this decrease is related to the adjustment factor for infinite number of plausible values used in the standard error evaluation formula. More importantly, it was found that the decrease in the estimated standard error with an increase in the number of plausible values used tend to move the estimated standard error further from the expected standard error values. Therefore, this may not reflect true situation in practice.

The empirical data studies were carried out using the NAPLAN 2015 Numeracy data. The results showed that there was no consistent pattern observed in the parameter estimates when the number of plausible values used was increased from 1 to 46. The decreasing trend in the estimated standard errors was not observed in the empirical data studies when the number of plausible values used increased from 3 to 46. This means that the number of plausible values used, up to 46, does not improve the estimates of NAPLAN parameters, nor the standard errors of the parameter estimates. However, when sample sizes were small ($N < 2,000$), the parameter estimates using one or three plausible values showed a large difference from the estimates using five or more plausible values. The simulation and empirical studies both show that the test reliability and sample size have a much larger effect on optimising the population parameter estimates than the effect of increasing the number of plausible values used.

Declaration

This is to certify that:

- I. this thesis comprises only my original work towards the PhD;
- II. due acknowledgement has been made in the text to all other material used; and
- III. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Candidate's Signature:

Acknowledgements

I would like to take this opportunity to express my thanks to the many people who have helped me complete this academic milestone.

Firstly, I would like to thank my supervisors. Professor Ray Adams and Dr Siek Toon Khoo have patiently supported and guided me. Without their persistence and endorsement, I would have never completed this difficult journey. I wish to sincerely thank them for sparing their valuable time teaching me and providing me advice.

Secondly, I would like to thank Dr Cuc Nguyen. Without your help as my supervisor I couldn't have successfully achieved this accreditation. Likewise, I wish to thank my committee chair Dr Zhonghua Zhang for his time and effort.

Additionally, I would also like to thank my colleagues in the Psychometric and Methodology program at the Australian Council *for* Educational Research, for their support and friendship. Particularly, I would like to thank my colleague and dear friend, Chris Rosicka, for her help and valuable comments. As well as, a special thanks to Dr Kate Reid and Dr Nathan Zoanetti for their help in proof-reading on my dissertation. I am very grateful for the support provided for this study by the Australian Council *for* Educational Research.

Finally, I would like to thank my husband Graeme and my darling daughter Renee for their understanding, support and tolerance throughout this time.

Table of Contents

List of Tables	iv
List of Figures	v
Chapter 1	Introduction..... 1
Chapter 2	Literature Review and Research Questions 11
2.1	Item Response Models 11
2.2	Estimating IRMs 17
2.3	Working with Plausible Values 25
2.4	The Initial Development of the Plausible Values Technnique 28
2.5	Where Have Plausible Values Been Used? 30
2.6	How Many Plausible Values Should Be Used for Secondary Analysis?..... 35
2.7	Research Questions 41
Chapter 3	Method 45
3.1	The Model Used in This Dissertation 45
3.2	Simulation Design..... 47
3.3	Identify the Number of Replications Used for Simulations..... 55
3.4	Data Generation 57
3.5	Evaluation of the Simulation Outcomes 59
3.5.1	Estimate Population Parameters 59
3.5.2	Evaluation of the Estimated Population Parameters..... 64
3.6	Illustration with the Empirical Data Studies..... 67

3.7	Summary	70
Chapter 4	Results of the Simulation Studies	71
4.1	Simulation One with Homogeneous Population.....	71
4.1.1	The Bias and RMSEs.....	72
4.1.2	The Estimated Standard Errors.....	84
4.2	Simulation Two with Heterogeneous Populations.....	91
4.2.1	The Bias and RMSEs of Regression Coefficient Estimates	93
4.2.2	The Bias and RMSEs of the Other Parameter Estimates.....	96
4.2.3	The Estimated Standard Error of Regression Coefficients.....	102
4.2.4	The Standard Errors of the Other Parameter Estimates.....	103
Chapter 5	Results of Empirical Data Studies	108
5.1	Results of Empirical Study One.....	110
5.2	Results of Empirical Study Two	123
Chapter 6	Discussion	135
6.1	The Summary of Findings by Each Study	135
6.2	Discussion by Research Questions	139
6.2.1	The Effect of Increasing Numbers of Plausible Values Used on the Population Parameter Estimates	139
6.2.2	The Effect of Increasing Numbers of Plausible Values Used on the Standard Error Estimates	139
6.2.3	The Effect of Test Reliability When Using Plausible Values	148
6.2.4	The Effect of Prior Distribution Assumption	153
6.2.5	The Effect of Sample Size	157

6.3	Should One Be Concerned with the Bias and RMSE?	157
Chapter 7	Conclusions and Recommendations	159
7.1	Conclusions.....	159
7.2	Recommendation for Future Work	162
References		164
Appendix 1	Item Parameters for Simulations Studies.....	172
Appendix 2	ConQuest Commands for Simulation Studies and Empirical.....	178
Appendix 3	Estimated Bias and RMSEs of Simulation One.....	185
Appendix 4	Standard Error Estimates of Simulation One.....	221
Appendix 5	Estimated Bias and RMSEs of Simulations Two	233
Appendix 6	Standard Error Estimates of Simulation Two	237
Appendix 7	The Results of Empirical Studies (NAPLAN 2015 Numeracy)	239
Appendix 8	Standard Errors of Mean Estimated by Three Methods.....	259

List of Tables

Table 2.1	Examples of Plausible Values Used in Large-scale assessments	35
Table 3.1	Design of the Simulation Studies.....	47
Table 3.2	Generating and Derived Parameter Values for Simulation One.....	61
Table 3.3	Generating and Derived Parameter Values for Simulation Two	61
Table 4.1	The Estimated Standard Errors of a Normally Distributed Sample ($N = 200$) with Different Length On-Target Tests, by Number of Plausible Values.....	85
Table 4.2	Generating Coefficient Values for Simulation Two	92
Table 4.3	Summary of the Simulation Results Relating to the Research Questions.....	107
Table 5.1	Test target Conditions by Year Level and by Jurisdictions and Australia.....	112
Table 6.1	Estimated Error Variance and Related Parameters.....	145
Table 6.2	Comparing the Standard Errors of the Mean When Estimated by Three Methods Using a Normally Distributed Sample ($N = 2,000$) with On-Target Tests	147
Table 6.3	Simple Statistics from the Generating Distribution and the Recovered Distributions.....	156

List of Figures

Figure 3.1	Test Target Conditions	50
Figure 3.2	Sample Distributions Used for Simulation One Data Generation	52
Figure 3.3	Sample Distribution Used for Simulation Two Data Generation	54
Figure 3.4	Illustration of Bias and RMSE	65
Figure 4.1	The Bias in the Parameter Estimates from Simulation One	78
Figure 4.2	The RMSEs of the Parameter Estimates from Simulation One	83
Figure 4.3	The Estimated Standard Error of the Parameter Estimates from Simulation One	90
Figure 4.4	The Bias in the Estimated Regression Coefficients from Simulation Two	95
Figure 4.5	The RMSEs of the Estimated Regression Coefficients from Simulation Two	96
Figure 4.6	The Bias in the Other Parameter Estimates from Simulation Two	99
Figure 4.7	The RMSEs of the Other Parameter Estimates from Simulation Two	101
Figure 4.8	The Standard Error of the Regression Coefficients from Simulation Two	103
Figure 4.9	The Standard Error of Five Parameters Estimates from Simulation Two	106
Figure 5.1	The Differences in the Mean Estimates in Empirical Study One	115
Figure 5.2	The Differences in the Standard Deviation Estimates in Empirical Study One	116
Figure 5.3	The Differences in the Percentage in the Top Band Estimates in Empirical Study One	117

Figure 5.4	The Differences in the 20 th Percentile Estimates in Empirical Study One	119
Figure 5.5	The Differences in the 80 th Percentile Estimates in Empirical Study One	120
Figure 5.6	The Standard Errors of the Parameter Estimates in Empirical Study One	123
Figure 5.7	The Differences in the Mean Estimates in Empirical Study Two	126
Figure 5.8	The Differences in the Variance Estimates in Empirical Study Two	127
Figure 5.9	The Differences in the Percentage in the Top Band Estimates in Empirical Study Two	129
Figure 5.10	The Differences in the 20 th Percentile Estimates in Empirical Study Two	130
Figure 5.11	The Differences in the 80 th Percentile Estimates in Empirical Study Two	131
Figure 5.12	The Standard Errors of the Parameter Estimates in Empirical Study Two	134
Figure 6.1	The Adjustment Factor by Number of Plausible Values	140
Figure 6.2	The Sampling, Measurement and Total Variances of the Mean Estimates from Simulation	142
Figure 6.3	The Sampling, Measurement and Total Variances of the Means from the NAPLAN Data	143
Figure 6.4	The Standard Errors of the Mean Estimated with Three Methods	148
Figure 6.5	The Likelihood Functions with Test Length Increases	151
Figure 6.6	Test Information Function and Estimated Standard Error with Three Test Lengths	152
Figure 6.7	Chi-squared ($df = 5$) Distribution	154

Figure 6.8	The Distribution of a Random Sample from the Original Chi-squared ($df = 5$) Distribution	154
Figure 6.9	Recovered Chi-squared ($df = 5$) Distributions	156

Chapter 1 Introduction

Large-scale assessments play an important role in evaluating and monitoring the outcomes of education systems. According to Lietz, Cresswell, Rust and Adams (2017), large-scale assessments are now being undertaken in about 70% of countries. Monitoring of outcomes of education systems are usually carried out through large scale assessment surveys that assess samples of students. Examples of international large-scale assessment surveys include international assessment programs such as Trends in International Mathematics and Science Study (TIMSS) (Martin & Kelly, 1996) and the Programme for International Student Assessment (PISA) (Adams & Wu, 2002). National monitoring of education systems could be done through large-scale assessment surveys such as the National Assessment of Educational Progress (NAEP) (Beaton, 1987) in the United States of America, and large-scale census assessment programs such as the National Assessment Program – Literacy and Numeracy (NAPLAN) (ACARA, 2015b) in Australia that assesses full cohorts of students in Years 3, 5, 7 and 9. The focus of large-scale assessment surveys is usually not on the performance of individual students¹, but rather on the performance of populations of students. Results from national large-scale assessments that involve full cohorts could focus on both the performance of individual students as well as on monitoring performance of the national population and subpopulations.

When the objective of an assessment program is on monitoring the performance of individuals, the focus of the assessment is on assessing domain proficiency of individual students. The aims are to optimise the measurement of the proficiency of each individual student, for certification purposes or for facilitating learning by identifying learning strengths and weaknesses to understand what a student can or cannot do, or for monitoring

¹ Throughout this dissertation, the term ‘student’ or ‘students’ is used to refer to subject or subjects to be measured. Alternative term: case, test taker, examinee or person, should all be seen as synonyms.

student learning progress over time. The assessment results can be useful for selection purposes, and for individual students and classroom teachers. The teachers can use assessment results to improve teaching and learning activities in the classroom.

In contrast to assessments designed to measure the performance of individuals, the focus of large-scale assessments is to accurately describe population characteristics. The aims of large-scale assessments are to provide optimal estimates of population parameters, to monitor population progress over time, to evaluate the quality of the educational system, and therefore to refine the education policies and strategies in an effort to improve the education system (Beaton and Barone, 2017; Lietz *et al.* 2017).

Education systems commonly cite the population parameter estimates derived from large-scale assessments as a way to communicate the effectiveness of their policies (or the shortcomings of past policies) to the public. Related commentary has often focused on changes in the mean estimated performance of student populations over time, or on differences between the mean estimated performance of different sub-populations within the assessed population (Thomson, De Bortoli and Underwood, 2017; OECD, 2017). To meaningfully determine whether such changes or differences represent real and statistically significant levels, in other words whether they were unlikely to have occurred by chance, the importance of accurately calculated parameters and accurately calculated standard errors for those parameters becomes clear. The statistical significance of such changes or differences can become a proxy for policy success or failure (OECD, 2017). Therefore, it is important to avoid inaccuracies in parameter estimates and corresponding standard errors that could lead to erroneous conclusions about changes in or differences between parameters. With the stakes so high for incumbent governments and system managers, there is a need to ensure that the accuracy and precision of estimation has been maximised within the given constraints. This points to the practical importance of the current study, which will focus on both parameter estimation accuracy and standard error estimation accuracy under different analysis conditions.

Beyond a focus on the mean estimated performance of populations or population sub-groups, education systems and commentators often refer to parameters relating to the spread of achievement of a given population or sub-population. Examples include examining the spread of scores within a population, either through statistics such as the variance, or alternatively through reporting the score range within which the middle 80% (e.g. NAPLAN) or 90% percent (e.g. PISA) of students are estimated to have scored. These can often be used to inform commentary about equity across systems or within a system over time (ACARA, 2015b; OECD, 2017; Thomson *et al.* 2017). Other population parameter estimates that may be cited by education systems include the percentage of students above or below some cut-point. These parameters might be commented on in relation to monitoring the effectiveness of education policies concerned with supporting less advantaged students to meet minimum proficiency standards (OECD, 2017), or at the other extreme, they may be cited in relation to monitoring the effectiveness of policies designed to extend high achieving students (e.g. State Government of Victoria, Australia, 2019). This breadth of population parameters that underpins policy development and debate reinforces the importance of research focusing on methodological factors that impact computational accuracy. Those charged with implementing large-scale educational assessments stand to benefit from a deeper understanding of how limitations in the data or modelling choices may degrade the accuracy of parameter estimation and secondary analysis.

Like physical measurements, educational assessment requires standardised instruments such as tests or surveys to determine the characteristics, ability, proficiency level, attitude or achievement of students or populations. An assessment instrument often consists of multiple questions or items and is administered to targeted students. The student item responses to the questions or to the items are then used to develop meaningful indicators through an appropriate model, to deduce a measurement indicator for each individual student (Stouffer, Guttman, Suchman, Lazarsfeld, Star & Clausen, 1950; Wilson 2005) or for estimating population characteristics. These measurements are then used for assessing educational outcomes either for an individual student or at a system level. The process of using a model to develop measurement indicators is called ‘scaling’. Scaling

enables comparison among students, schools or systems, as well as allowing comparisons over time. Most commonly used methodology for carrying out the scaling process, especially in large-scale assessments, is based on Item Response Models (IRMs)² (Berezner & Adams, 2017).

IRMs are based on the idea that the probability of a student successfully answering a test item can be modelled by a characteristic, referred to as the ability³ or proficiency, of the student. There are many different models that have been developed based on IRMs since the last century. A brief history of IRM is provided in Chapter 2. IRMs were initially developed to facilitate individual student ability estimation. This is one type of IRM that is concerned with assigning accurate measures to individual students. An optimal individual measurement can be achieved by increasing test length with the test at the appropriate ability level for the students (Lord, 1959; Gardner, 1970). Large-scale assessments are often required to cover a broad learning area using a large number of test items. However, a long test to be used in large-scale assessment can be expensive and difficult, as well as cause fatigue to students. It is desirable that large-scale assessments are only administered to a sample of students, and that each participating student is administered a small fraction of items from a large item pool due to financial and/or time constraints, as well as an intention to reduce the burden of a long test on individual students. The large-scale assessment surveys mentioned before are sample based tests and use a rotated test design (Beaton, 1987; Mislevy, Beaton, Kaplan, & Sheehan, 1992a; Martin & Kelly, 1996; Adams & Wu, 2002) with overlapping test booklets that are relatively shorter than those designed for optimal individual assessments.

² Many people use the noun phrase item response theory (IRT), others use the noun phrase item response model (IRM). In this dissertation, the term IRM is used throughout.

³ The term 'ability' is used throughout this dissertation to refer to the latent variable that is being measured, unless a special term was used in an original paper. Ability is therefore a synonym for other terms such as latent trait ability, student proficiency, case parameter or case estimates that are used throughout the literature.

A population distribution is generally conceptualised or assumed to be a continuous distribution. Although the population parameters can be estimated using individual students estimates, the distributions constructed using individual ability estimates are often discrete. It is difficult to make inferences from discrete distributions to be consistent with the underlying continuous distribution. The issue of estimating continuous distributions from discrete distributions is severely compromised when a test is short. A short test will not only compromise the accuracy of student ability estimates, but will also result in biased estimations of population parameters using individual ability estimates (Mislevy, 1985; Adams, Wilson & Wu, 1997a; Wu, 2005). It has been shown (Mislevy, Johnson & Muraki, 1992b) that even when a test is long enough and the estimated ability is a consistent estimate for each student, the sample variance of the estimated student abilities is inconsistent and an overestimate of the population variance. It has been pointed out that an appropriate way to estimate population parameters is to base the estimates directly upon the student response data rather than to obtain them from derived data, such as individual ability estimates (Mislevy, 1984).

The desire to directly estimate population parameters has prompted the development of another type of IRM. This type of IRMs provides a means of directly estimating population parameters without firstly estimating individual parameters. This type of model combines conditional IRM with a population density function so that it becomes an unconditional probability model. The model parameters (population parameters) are estimated using the marginal maximum likelihood (MML) technique (Bock and Lieberman, 1970). Since the population parameters are directly estimated from the model, it overcomes the shortcomings of using individual ability estimations, especially from short tests, to estimate population parameters. This has made it possible to use short tests in large-scale assessments to estimate population parameters where the student parameters are not a concern. For example, a student who participated in NAEP tests could be administered as few as two items (Beaton, 1987).

Although population parameters are estimated directly from this type of IRM model, a so-called plausible values (PVs)⁴ technique was developed for analysing NAEP 1983-84 results (Mislevy *et al.* 1992a). Plausible values are generated and provided for the convenience for secondary analyses and also to handle the uncertainty due to the complex sample design (Mislevy *et al.* 1992b). The plausible values for each student are random draws from the student posterior distribution (more detail is given in Chapter 2). Since student ability is not directly observed, it is a latent variable (or latent ability) that could be cast as missing value. Inference about it can be dealt with by an implementation of Rubin's (1987) multiple imputation technique for dealing with missing data. Rubin (1987) demonstrated that a single imputation for missing data performs poorly and possibly underestimates the variance. A single imputation assumes that the imputed value is the true value of the missing data. A better way to impute for missing data is to use multiple imputations with an appropriate specified model. Because the imputations, or plausible values, are randomly drawn from a distribution, the estimates from plausible values include random variation. Multiple imputations will result in different estimates when different number of plausible values are used especially when the number of plausible values is small.

With the multiple imputation technique, multiple plausible values are imputed for each student sampled from the underlying population. The uncertainty of student ability estimates due to unobserved latent ability is captured from the variation among the multiple plausible values. Multiple plausible values are generated for all students and each resulting set of generated data over students can be considered as a complete dataset. Each of these datasets is then analysed separately and the final population parameters are estimated by averaging the parameters obtained from the multiple datasets. The plausible value technique is reviewed in greater details in Chapter 2. This dissertation is concerned

⁴ The term 'plausible value' or 'plausible values' are used in main text of this dissertation, the abbreviation 'PV' or 'PVs' are used in tables and graphs.

with the contexts in which the estimation of population parameters, such as the population mean and variance, is the primary focus.

Since the NAEP started utilising the plausible value methodology in the 1980s, plausible value technique has been adopted in many large-scale assessments such as TIMSS (Adams, Wu, & Macaskill, 1997b), PIRLS (Martin, Mullis, Gonzalez, & Kennedy, 2003), PISA (Adams & Wu, 2002) and NAPLAN (ACER, 2010). NAPLAN assessments are administered to all students at targeted year levels rather than to a sample of students. The focus of NAPLAN is both on reporting individual performance to schools and parents and on monitoring the educational systems for the population. Plausible values are used for estimating population parameters.

The number of imputations required has been a significant question raised by researchers who use the multiple imputation technique to deal with missing data. Much research has been conducted on multiple imputations in dealing with missing observed values (Little & Rubin, 1983; Schafer, 1997; Allison, 2002). In this context, five imputations were usually considered sufficient for all settings. However, the research on the number of plausible values that should be used in an IRM context is limited. The number of plausible values for accurate estimation of population parameters is often assumed, but there is no empirical evidence supporting these assertions. For instance, Luo and Dimitrov (2019) assert that it is 'well-known' that five plausible values are sufficient; however, the source of these claims is not clear. A small simulation study by Wu (2005) showed that five plausible values recover the generating values well, in terms of the population mean and variance. Research on the advantages, if any, of using more than five plausible values is practically non-existent. In the majority of large-scale assessments and surveys, five plausible values have been generated for each student to estimate population parameters, including population means, variance, standard errors, percentiles, percentages in a defined proficiency band and regression coefficients. In recent years, however, some of the large-scale assessment surveys have increased the number of plausible values they used to well beyond five imputations (National Assessment of Educational Progress, 2016; OECD, 2013; OECD, 2017). For example,

ten plausible values were used for PISA 2015 and PISA 2018 (only five plausible values were used for PISA 2000 to PISA 2012) in an attempt to improve the accuracy of the estimated standard error (OECD, 2018) and 20 plausible values were used in recent NAEP assessments (National Assessment of Educational Progress, 2016). Increasing the number of plausible values used increases the time required for secondary analysis, which would also increase the data storage cost, especially when the sample size is large. Increasing the number of plausible values used to provide improved accuracy in the population parameter estimates needs to be justified. There has not been published research that clearly establishes the justification for increasing the number of plausible values used beyond five in IRM context. The objective of this dissertation is to investigate the number of plausible values that are required, under different assessment and research conditions, for producing population estimates with acceptable accuracy and precision.

A number of factors which can possibly affect the outcomes in the number of plausible values used in estimating population parameters have been identified for investigations. These include sample size, test length, test targeting conditions and prior sample distribution assumptions. The research focus of this dissertation is in the relationship between the number of plausible values used in the IRM context and the properties of population parameter estimates. That is, how increases in the number of plausible values used will affect population parameter estimates. Also of interest are those factors that may affect the estimates of population parameters produced using plausible values. They are: test length; test targeting condition; sample distribution and sample size. Other research questions are: Will increasing the number of plausible values used reduce the standard error of estimated parameters? Will the number of plausible values required vary depending on population distributions, sample sizes and test characteristics to provide similar accuracy of the population parameter estimates?

These research questions warrant consideration as there is no systematic research can be found that is investigating how many plausible values should be used in an IRM context. The current widely accepted practice of using five plausible values for providing optimal

estimates of population parameters may be adequate. On the other hand, it may be necessary to use fewer than or more than five plausible values. In an attempt to answer these research questions, two simulation studies and two empirical data studies were carried out as listed below:

Simulation One: The first simulation study was conducted over a range of sampling and test conditions to investigate the relationship between the number of plausible values used and the population parameter estimates. In this simulation study, the sampling and test conditions were simulated to investigate the effect of increasing the number of plausible values used on the population parameter estimates. Various sample sizes, population distributions and test conditions, in terms of test length and test targeting status, were investigated. There were no population covariates incorporated in the data generation. This is based on the assumption that the students are sampled from a homogeneous population and have similar characteristics.

Simulation Two: The second simulation study was carried out using a reduced set of simulation conditions based on the findings of the first simulation study. Student background covariates were additionally incorporated in generating simulation data and were used as conditional variables for drawing plausible values. This means that the generating samples are from a heterogeneous population, and students may have different characteristics depending on their background.

Empirical Study One: The findings of the simulation studies were further explored with empirical data. The empirical data studies were carried out using the NAPLAN 2015 Numeracy data across four year levels, Years 3, 5, 7 and 9. In Empirical Study One, the results were evaluated with data from all participating students.

Empirical Study Two: The second empirical data study was carried out with a much smaller subset of data than the data used in Empirical Study One. It is based only on the data from Indigenous students to study the effects on small samples.

These two empirical data studies were aimed at evaluating the effect of increasing the number of plausible values on the NAPLAN population parameter estimation.

This dissertation consists of seven chapters. Following this introduction chapter, a brief history of IRMs is presented and two types of IRM models are discussed in Chapter 2. Chapter 2 includes a literature review on the use of plausible values in the IRM context, the current practice on the number of plausible values used in large-scale assessments, and the research questions for this dissertation.

Chapter 3 describes the methodology used in this dissertation which includes the simulation design, the data generation procedure, the indices used for evaluating the simulation studies, and the method used for the empirical data studies. The method of how to work with plausible values is also described in this chapter. In addition, this chapter investigates how many replications is sufficient for evaluating the outcomes of the simulation studies.

Chapter 4 reports the results from the two simulation studies based on data generating models with and without student background variables as conditioning variables. The results from Simulation One are given in the first part of Chapter 4. The second part of Chapter 4 describes the results from Simulation Two using a data generating model with student background variables as conditional variables.

Chapter 5 presents the results from the empirical data studies. The results from Empirical Study One are presented first, followed by the results from Empirical Study Two.

Chapter 6 discusses the outcomes of the simulation studies and the outcomes from the empirical data studies, as well as describing how the findings relate to the research questions. The implications of the findings for current practice are also discussed. Finally, conclusions based on this dissertation are provided in Chapter 7, with some recommendations for possible future work.

Chapter 2 Literature Review and Research Questions

This chapter consists of seven sections. Section 2.1 describes item response models with a brief review of their history. Section 2.2 reviews the estimation of IRMs and the generation of plausible values. Section 2.3 describes the procedures used in working with plausible values. Section 2.4 describes the reasons for using plausible values, and the advantages and disadvantages of using plausible values to estimate population parameters through secondary analysis. This is then followed by Section 2.5 with a review of the applications of plausible values used in large-scale assessments and surveys in the education field. Section 2.6 reviews some of the previous research on the number of plausible values used in population parameter estimates. Finally, Section 2.7 describes the research questions this study proposes to answer.

2.1 Item Response Models

IRMs have been widely applied in psycho-social science fields since early last century. The fundamental statistical idea underlying IRMs is that the probability of a student responding to an item is a function of a student's location on a latent trait (referred to as student ability or proficiency) and a set of item characteristics (referred to as item parameters). Student abilities are interpreted as locations on a continuous metric (scale)⁵. Item response models are mathematical representations of the probability of a correct response as a function of student ability and item parameters, such as item location parameter, item discrimination parameter and/or guessing parameter (van der Linden,

⁵ The models considered here are those that deal with continuous latent traits. Latent class models are not considered.

2016). One of the most important outcomes of IRMs is that the item location parameters and student abilities are placed along the same underlying continuous latent metric.

The foundations of item response theory were laid out in a journal article by Alfred Binet in 1904 (van der Linden, 2016). Binet was the first person to introduce the idea of scaling, to map both items and students on a single scale. He first estimated the scale value of items using empirical data, then scored students using the scale values. The importance of Binet's work was not recognised until 1925. In 1925, Thurstone (1925) published a paper which re-examined and plotted the proportions of children in successive age-groups who correctly answered a given test question in Binet's original test. Thurstone assumed an unknown latent scale for items, then estimated the locations of all test items on an age scale with an arbitrary origin (average ability of children aged 3.5 years) and unit (the standard deviation of the children aged 3.5 years). By doing so, Thurstone effectively disengaged the intelligence latent variable from age in his scale. His plot and scaling method (so-called absolute scaling method) related chronological age to the probability of success on a given item, describing the probability of successfully answering an item as a function of ability.

In 1950, Lazarsfeld wrote a book chapter (1950) that analysed the dichotomously classified responses⁶ of persons who have been administered multiple-item attitude questionnaires. In this chapter, Lazarsfeld assumed that a one-dimensional continuum (a person's attribute) existed. Based on the response pattern of the administered items, people can be ranked along this continuum. The probability that respondents provided a "positive answer" for the item i is a function of their position on the continuum. Lazarsfeld plotted the probability against the continuum, he called the function line the "trace line of item i ".

⁶ dichotomously classified or dichotomously scored item refer to an item response has only two possible scores (Correct/Incorrect, True/False or Yes/No, etc.).

Latent Trait Theory assumes that the test taker's performance on a test can be described by the test takers' characteristics. As these traits are not directly measurable and are unobserved they are referred to as latent traits. Lord (1953) in his paper on relating test scores to latent traits also assumed that test takers can be ordered on a single dimension. Lord plotted the proportion of correct answers along the latent trait continuum following Tucker's (1946) terminology and called this the Item Characteristic Curve (ICC). Unlike Thurstone's plots, the ability continuum in Lord's plots is not directly measured. The metric used for representing ability is arbitrarily chosen. The Latent Trait Theory that Lord was talking about is now commonly referred to as Item Response Theory.

In the late 1950s, Birnbaum introduced the term Item Characteristic Curve models and presented his work on a 3-parameter logistic model (Bock, 1997). The Item Characteristic Curve models the relationship between a student's latent ability and the probability of the student correctly responding to an item. However, Birnbaum's work was not well known until 1968, when Lord and Novick incorporated the Birnbaum 3-parameter logistic model in their book, *Statistical Theories of Mental Test Scores* (1968). A 3-parameter IRM model (Birnbaum 1968) for a dichotomously scored items was expressed as:

$$P(X_{ni} = 1; \theta_n, \delta_i, a_i, c_i) = c_i + (1 - c_i) \frac{\exp\{a_i(\theta_n - \delta_i)\}}{1 + \exp\{a_i(\theta_n - \delta_i)\}} \quad \text{Equation 2.1}$$

Where: X_{ni} represents the response of student n to item i , 1 if correct and 0 if incorrect

θ_n is referred to as the ability parameter for student n

δ_i is referred to as the item parameter for item i

a_i is the discrimination parameter of item i

c_i is the guessing parameter of item i

Parallel to Lord's studies, Georg Rasch, a Danish mathematician who had been working with latent trait measurements since the 1940s, developed the well-known Rasch Model in 1960 (Rasch, 1960/1980). This is often referred to as the 1-parameter logistic (1PL) model:

$$P(X_{ni} = 1; \delta_i, \theta_n) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \quad \text{Equation 2.2}$$

It is noted that when $c_i = 0$, Equation 2.1 becomes a 2-parameter logistic IRM model and when $c_i = 0$ and $a_i = 1$, Equation 2.1 becomes a 1-parameter simple logistic IRM model as noted in Equation 2.2.

Both Equation 2.1 and Equation 2.2 model the probability of response X_{ni} from student n to a particular item i of a test paper.

The introduction of the concept of local independence provided a crucial step in the development of IRMs (Lazarsfeld, 1950). Local independence means that after controlling for a person's latent ability, the response to any item in a test is unrelated to the response for any other items. This implies that there is no relationship among items after a person's latent abilities are controlled for, although the responses to items may be inter-correlated for the whole sample. With this assumption, the response pattern likelihood is the simple product of the individual item likelihoods. The joint likelihood from N students responding to I items can be expressed as below:

$$\Lambda(\theta, \xi; x) = \prod_n^N \prod_i^I P(x_{ni}; \delta_i, \theta_n) \quad \text{Equation 2.3}$$

where ξ is a vector of I item parameters and x is a vector of responses from N students to I items.

Many IRM models have been developed to serve different purposes, for example the Rasch model (1-parameter IRM model) models the probability of a student responding to an item using just a location parameter; the 2-parameter item response model adds a parameter to describe items with different discriminations (Birnbbaum, 1968); the 3-parameter model adds a guessing parameter (Birnbbaum, 1968; Lord, 1980). Other developments allowed for items with multiple response categories (the rating scale model,

Andrich, 1978; and the partial credit model, Masters, 1982), multiple dimensions (Reckase, 2009) and multiple facets (Linacre, 1993).

As mentioned in Chapter 1, it is useful to classify IRMs into two types, developed to serve two broad types of application: one focussed on estimating individuals' proficiency and a second focussed on estimating population parameters. Equation 2.1 to Equation 2.3 belong to the first type of IRMs that are used to obtain individual ability estimates. In these models the parameters to be estimated are student ability θ_h and item parameter δ_i and no population parameters are included; both student ability and item parameter are point estimates.

The primary aim of large-scale assessments is to provide information on educational outcomes at a system level and to describe the population's characteristics, rather than to provide information about individual student learning outcomes. Point estimates of individual student parameters (θ_h) from the first type of IRM are optimal for reporting individual proficiency (Mislevy *et al.* 1992b) with test length that is sufficiently long. However, point estimates could have a distribution that is seriously biased as an estimate of the θ distribution for the population (Little & Rubin, 1983). In the Rasch modelling, a student's total raw score on a test is a sufficient statistic for the ability estimates (Andersen, 1973). This means that student ability estimates will be based on the total raw score, and students with the same total raw score will have the same ability estimates even if their response patterns are different. As an example, a 20-item test with raw scores from 0 to of 20, would only have 21 discrete ability estimates produced using the Rasch model: one estimate per raw score point, regardless of the number of students. When using 21 discrete point- estimates to construct a population distribution, it would result in a discrete distribution. This discrete distribution may result in biased population estimations, e.g. a biased estimation of the proportion of students in a particular proficiency level (Wu, 2005; Berezner & Adams, 2017). The problem of discrete student ability estimates may reduce with the 2-parameter IRM model, where there is an ability estimate for each response pattern providing the discriminations of items are different. However, when a test is short,

the number of response patterns is relatively small, so it will still result in a discrete population distribution.

The problem of using individual point ability estimates to estimate population parameters is exacerbated in large-scale assessments. In order to cover a broad range of teaching content, a large number of items are required to be administered, which could result in very long tests. To reduce the burden of a long test on individual students, as well as the cost, most large-scale assessments implement a multiple-matrix item sampling design. With this design, all of the tested items are divided into multiple clusters, each cluster only contains a small number of items, and they may or may not have overlapping items across clusters (Mislevy *et al.* 1992a). Each sampled student would be administered a subset of tested items consisting of items from one or more clusters (Berezner and Adams, 2017).

The issues of using the point ability estimates, especially from a short test, to estimate population parameter as described above was overcome with the second type of IRM. The first type of IRMs (Equation 2.3) is a conditional model, that is, the probability of response pattern x_n from student n is conditional on the student latent ability θ_n . To transform the conditional model to an unconditional model, a population density $f_\theta(\theta|\alpha)$ is combined with the conditional model as below:

$$f(x_n; \xi, \alpha) = \int_{\theta} f(x_n; \xi | \theta) f_{\theta}(\theta | \alpha) d\theta \quad \text{Equation 2.4}$$

In Equation 2.4, the function $f(x_n; \xi | \theta)$ is an IRM model, and $f_{\theta}(\theta | \alpha)$ is a population model which specifies the density of the latent variable θ , α is a vector of parameters of the distribution, for example the mean and variance of the distribution or observed students background response (Bock & Aitkin, 1981). Under this model, the student latent ability variable θ is not estimated but sampled from the specified population and population parameters (α) were estimated directly without first estimating individual parameters.

2.2 Estimating IRMs

In 1925, Fisher (1925) introduced the maximum likelihood (ML) method as an approach for estimating model parameters. Since then the maximum likelihood method has become the predominant approach in IRMs over other techniques such as the ordinary least squares or the minimum chi-square method and others that can be used to estimate the student latent ability.

Maximum Likelihood Estimation (MLE) for IRMs was made easier with two developments. First, the introduction of the logistic item response function as an alternative to the normal ogive model by Fisher and Yates (1938); and second, Lazarsfeld's (1950) application of the concept of local independence. These developments allowed for the joint maximum likelihood (JML) method to be developed by Birnbaum (1968), and Wright and Panchapakesan (1969). With JML, all items and individual abilities are estimated jointly by maximising the likelihood specified in Equation 2.3. The JML provides point estimates for item parameters and student abilities simultaneously and has been widely used (Baker, 1992) for estimating individual proficiency. As mentioned in the previous section, the model is specified as a conditional IRM where the probability of the response vector x is conditioned on a given latent ability. In this model (Equation 2.3), there is a set of N unknowns which consists of individual student ability (θ_n), and a vector that consists of item parameters (δ_i) which are also unknown.

Under the JML method all item parameters (δ_i) and all person abilities (θ_n) are regarded as unknown fixed parameters and estimated simultaneously. Neyman and Scott (1948) have shown that the MLEs need not be consistent when all parameters are estimated together with an increasing sample size. Andersen (1973) showed that the JML

estimation of item parameters for the Rasch model are not consistent when the number of items is fixed but the sample size approaches infinity⁷.

Bock and Lieberman (1970) introduced the assumption of a population distribution based on a random effect as the means to address the problem of inconsistency. This assumption change can be seen as a change in the specification of the model and the estimation is called marginal maximum likelihood (MML). The assumption changed the ability θ from an unknown parameter in Equation 2.3 to a random effect in the model specified in Equation 2.4. Under this model, the student latent ability variable θ is not estimated but rather sampled from a specified population. That is, given a particular response pattern x from a student who is sampled from a population distribution density $f_{\theta}(\theta | \alpha)$, the likelihood of the parameters α, ξ using X is maximised in MML over N sampled students:

$$\Lambda(\alpha, \xi; X) = \prod_{n=1}^N \int_{\theta} f(x_n; \xi | \theta) f_{\theta}(\theta | \alpha) d\theta \quad \text{Equation 2.5}$$

The advantage of MML is that if both the IRMs and the assumed population distribution are correct, the item parameter estimates from MML are consistent for any fixed number of items (Bock & Aitkin, 1981). When interest is in population parameters this approach avoids the issues confronted when using the first type of model that includes individual student parameters and requires a 2-step process for estimating population parameters: estimating individual ability first and then estimating population parameters.

Although not a requirement, $f_{\theta}(\theta | \alpha)$ is often assumed to be the Gaussian density – that is, that the students have been sampled from a normally distributed population with mean

⁷ Conditional maximum likelihood provides consistent estimates if sufficient statistics are available, but only for the Rasch model; it is not considered further in this study.

μ and variance σ^2 (Bock & Lieberman, 1970; Bock & Aitkin, 1981; Thissen, 1982; Mislevy, 1984; Adams & Wu, 2007) for a practical reason. That is:

$$f_{\theta}(\theta; \alpha) \equiv f_{\theta}(\theta; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2}\right] \quad \text{Equation 2.6}$$

or, equivalently,

$$\theta = \mu + E \quad \text{Equation 2.7}$$

where $E \sim N(0, \sigma^2)$

Adams *et al.* (1997a) have discussed, how a natural extension to Equation 2.7 is to replace the mean with the regression model, $Y_n^T \beta$, where Y_n is a vector of μ , fixed and known values for student n , and β is the corresponding vector of regression coefficients, such as gender or language background status. The population model for student n then becomes:

$$\theta_n = Y_n^T \beta + E_n \quad \text{Equation 2.8}$$

where E_n is independent and identically distributed with $N(0, \sigma^2)$.

Hence the population model Equation 2.7 becomes:

$$f_{\theta}(\theta_n | Y_n, \beta, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\theta_n - Y_n^T \beta)^T (\theta_n - Y_n^T \beta)\right] \quad \text{Equation 2.9}$$

which is a normal distribution with mean $Y_n^T \beta$ and variance of σ^2 . When Equation 2.9 is used as the population model in Equation 2.4, then the parameters estimated are β , σ^2 and ξ .

For N sampled students, the likelihood is:

$$\Lambda(\xi, \beta, \sigma^2 | X) = \prod_{n=1}^N \int_{\theta} f(x_n; \xi | \theta) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2) d\theta \quad \text{Equation 2.10}$$

and the log likelihood is:

$$\lambda(\xi, \beta, \sigma^2 | X) = \sum_{n=1}^N \log \int_{\theta} f(x_n; \xi | \theta) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2) d\theta \quad \text{Equation 2.11}$$

Equation 2.10 (or Equation 2.11) shows that this model includes parameters (β, σ^2) of the distribution $f_{\theta}(\theta; Y_n, \beta, \sigma^2)$ together with item parameters ξ , but no individual ability parameters (θ) are included or estimated. Estimates of the parameters are obtained by maximising the log-likelihood (Equation 2.11).

Differentiating $\lambda(\xi, \beta, \sigma^2 | X)$ with respect to each of the parameters and defining the marginal posterior as:

$$h_{\theta}(\theta_n; Y_n, \beta, \sigma^2, \xi | x_n) = \frac{f_x(x_n; \xi | \theta_n) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2)}{f_x(x_n; Y_n, \beta, \sigma^2, \xi)} = \frac{f_x(x_n; \xi | \theta_n) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2)}{\int_{\theta} f_x(x_n; \xi | \theta_n) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2)} \quad \text{Equation 2.12}$$

the following series of likelihood equations are derived (Adams et al. 1997a):

with respect to ξ

$$A' \sum_{n=1}^N \left[x_n - \int_{\theta_n} E_z(z | \theta_n) h_{\theta}(\theta_n; Y_n, \xi, \beta, \sigma^2 | x_n) d\theta_n \right] = 0 \quad \text{Equation 2.13}$$

where

$$E_z(z | \theta_n) = \Psi(\theta_n, \xi) \sum_{z \in \Omega} z \exp[z'(b\theta + A\xi)] \quad \text{Equation 2.14}$$

with respect to β

$$\sum_{n=1}^N \int_{\theta_n} (Y_n \theta_n - Y_n Y_n' \beta) h_{\theta}(\theta_n; Y_n, \xi, \beta, \sigma^2 | x_n) d\theta_n = 0 \quad \text{Equation 2.15}$$

$$\hat{\beta} = \left(\sum_{n=1}^N Y_n Y_n' \right)^{-1} \left(\sum_{n=1}^N \bar{\theta}_n Y_n \right) \quad \text{Equation 2.16}$$

where

$$\bar{\theta}_n = \int_{\theta} \theta_n h_{\theta}(\theta_n; Y_n, \xi, \beta, \sigma^2 | x_n) d\theta \quad \text{Equation 2.17}$$

with respect to σ^2

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N \int_{\theta} (\theta_n - Y_n' \beta)^2 h_{\theta}(\theta_n; Y_n, \xi, \beta, \sigma^2 | x_n) d\theta \quad \text{Equation 2.18}$$

These equations can be solved using an EM algorithm (Dempster, Laird & Rubin, 1977) following the approach of Bock and Aitkin (1981).

Unfortunately, the integrations in MML (Equation 2.13 to Equation 2.17) can be very complex and difficult to compute and are typically numerically approximated with either quadrature or Monte-Carlo methods. Equation 2.4 can be approximated as:

$$f(x_n; \xi, Y_n, \beta, \sigma^2) \approx \sum_{q=1}^Q f(X; \xi | \Theta_q) W_q(\Theta_q; Y_n, \beta, \sigma^2) \quad \text{Equation 2.19}$$

and the marginal posterior Equation 2.12 can be approximated as:

$$h_{\Theta}(\Theta_q; Y_n, \beta, \sigma^2, \xi | x_n) = \frac{f_x(x_n; \xi | \Theta_q) W_q(\Theta_q; Y_n, \beta, \sigma^2)}{\sum_{q=1}^Q f_x(x_n; \xi | \Theta_q) W_q(\Theta_q; Y_n, \beta, \sigma^2)} \quad \text{Equation 2.20}$$

where Θ_q is referred to as a quadrature point or a node, and $W_q(\Theta_q; Y_n, \beta, \sigma^2)$ is the weight defined for each node.

Three popular integration methods used in this context are the Bock-Aitken quadrature, the Gauss-Hermite quadrature and Monte-Carlo integration. The Bock-Aitken quadrature method is a naïve procedure in which a uniform grid of nodes ($\Theta_q, q = 1, 2, \dots, Q$) are chosen, and at each grid point, weight of the node (W_q) is defined by the function $f(\theta | \alpha)$. Thus, the sum of ($\Theta_q * W_q$) is the approximation of the sum of the area defined by the function $f(\theta | \alpha)$. When the grid points are increased, the approximation becomes more accurate.

The difference between the two quadrature methods is in the way the grid points and weights are prepared. The grid points in the Bock-Aitken quadrature are fixed and equally distanced and the weight is defined by the function $f(\theta | \alpha)$. The Gauss-Hermite quadrature is conceptually similar to the Bock-Aitken quadrature with the grid points and weights being chosen to optimise the approximation. That is, the grid points are given by the roots of Hermite polynomials and the weight is determined with the weighting function $W_q = e^{q^2}$ (Abramowitz & Stegun, 1972, p. 890). With the Gauss-Hermite quadrature, if there is a little density at the extreme ends of the curve, the grid points are chosen further apart, which gives them less weight in the approximation; and at the middle part of the curve which has higher density, grid points will be closer to each other and given more weight in the approximation. The third approach is a Monte-Carlo method in which the grid points $\Theta_q, q = 1, 2, \dots, Q$, are random draws in a unit area and the weight of the grid points are all equal one. A grid point is selected from the area below the function $f(\theta | \alpha)$. The frequency of these points selected in the Monte-Carlo method is the self-weighting of $1/Q$. The number of the selected grid points over total grid points is the approximation of the sum of the area defined by the function $f(\theta | \alpha)$.

There are other methods developed to approximate the integrals. For example, adaptive Gaussian quadrature and Laplace approximations. In brief, adaptive Gaussian quadrature approximates the integral of a function $f(\theta|\alpha)$ using static quadrature rules on adaptively refined subintervals of the integration domain (Press, Teukolsky, Vetterling & Flannery, 2007). Laplace approximation is a method that uses a normal distribution to approximate a function $f(\theta|\alpha)$. It first locates the mode of a function $f(\theta|\alpha)$, takes this as the mean of the normal approximation, and then calculates the variance of the normal distribution by examining the curvature of the function $f(\theta|\alpha)$ at the mode. Laplace approximation is likely to work well when the function $f(\theta|\alpha)$ is unimodal and relatively symmetric around the mode (Peng, 2018). Although adaptive Gaussian quadrature and Laplace approximations can lead to moderate increases in the speed for integration, they do not address the fundamental problem when there are a large number of dimensions. Neither of these methods scale a large number of dimensions very well (Pinheiro & Chao, 2012). Therefore, they are moderate improvements on the ordinary quadrature method, but they are not an alternative to Monte Carlo integration. For complex models, a more promising development is the Markov Chain Monte Carlo (MCMC) (van Ravenzwaaij, Cassey & Brown, 2018) method. This has become a popular method, but it has not yet been used in the context of large scale assessments, which is the topic this dissertation. This is an interesting area that is worthy of further investigation.

To estimate population parameters specified in Equation 2.15 to Equation 2.18, the integrations can be approximated by Equation 2.19 and Equation 2.20 by one of many integration methods. When the Monte-Carlo method is used, a random sample must be drawn from the student posterior distribution that is specified by Equation 2.20. These randomly drawn samples (points θ_q), with the Monte-Carlo method from Equation 2.20, are the plausible values (Volodin & Adams, 1995; Monseur & Adams, 2009).

The following is a description of the approach Volodin and Adams (1995) developed for drawing the plausible values and is the method used for this dissertation:

First draw M vector-valued random deviates, $\{\varphi_{mn}\}_{m=1}^M$, from the multivariate normal distribution (prior distribution as in Equation 2.9), $f_{\theta}(\theta_n; Y_n, \beta, \sigma^2)$, for each student n . These vectors are used to compute an approximation for the integration in the posterior distribution Equation 2.12, using Monte-Carlo integration:

$$\int_{\theta} f_x(x_n; \xi | \theta_n) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2) d\theta \approx \sum_{m=1}^M f_x(x; \xi | \varphi_{mn}) \equiv \mathfrak{I} \quad \text{Equation 2.21}$$

At the same time, the values

$$p_{mn} = f_x(x_n; \xi | \varphi_{mn}) f_{\theta}(\varphi_{mn}; Y_n, \beta, \sigma^2) \quad \text{Equation 2.22}$$

are calculated, hence a set (M) of pairs $(\varphi_{mn}, p_{mn} / \mathfrak{I})_{m=1}^M$ are obtained, which can be used as an approximation of the posterior density Equation 2.12; and the probability that φ_{mn} could be drawn from this density is given by:

$$q_{mn} = \frac{p_{mn}}{\sum_{m=1}^M p_{mn}} \quad \text{Equation 2.23}$$

At this point, $L = M * N$ uniform distributed random numbers, $\{\eta_i\}_{i=1}^L$ are generated; and for each random draw, the vector φ_{mn} that satisfies the condition:

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_i \leq \sum_{s=1}^{i_0} q_{sn} \quad \text{Equation 2.24}$$

is selected as plausible vector.

It is worth noting that a plausible value is not an estimate of a student's ability or an individual student's proficiency score, rather plausible values are randomly drawn from

a student's posterior distribution $h(\theta|x)$. In other words, instead of estimating a student's ability, a probability distribution for a student's ability is estimated. According to Mislevy *et al.* (1992a), for a given individual, a plausible value is neither unbiased nor efficient, nor does it minimise mean square error over the population estimates. When there is more than one plausible value drawn, the multiple plausible values not only provide information about a student's ability distribution estimates, but also the uncertainty associated with the estimates.

Because the interest of this dissertation is in the population parameter estimates, the second type of IRM model specified in Equation 2.4 was used throughout this dissertation. The items modelled in this dissertation are all dichotomously scored. The item response model used is a simple logistic model implemented as a special case of the Mixed Coefficient Multinomial Logit model (Adams, Wilson & Wang 1997c) with ACER ConQuest 4 (Adams, Wilson & Wu, 2015). It is not expected that this choice would limit the generality of the findings, but as discussed in chapter 7, further investigations with alternative models would be of value .

2.3 Working with Plausible Values

Plausible values can be seen from two aspects: In the IRM model estimation, the Monte-Carlo integration is used to solve the MML equations. Mathematically, the grids used in the Monte-Carlo integration are actually the plausible values. On the other hand, when student parameters are cast as missing data, plausible values are multiple imputations for the missing latent proficiency of each student, based on the student posterior distribution. These multiple imputations can be stored as latent proficiency data to be provided for secondary data analyses. Data with plausible values reflecting latent proficiency make these secondary analyses possible using standard statistical tools.

Plausible values are used as an intermediate step for estimating population characteristics and for secondary data analysis, for example: for estimating subpopulation statistics, for

estimating proportions of students within a proficiency level, and for estimating the relationship between proficiency and various social and educational variables.

As described in the previous section, the population parameters can be estimated directly from the second type of IRM model. However, researchers are often not only interested in population parameter estimates, but are also interested in sub-population characteristics, in making comparisons across populations, in monitoring population growth over time and in statistical modelling of data. It is important to know that population parameters are not estimated from plausible values, rather plausible values reflect the estimates of population parameters that were obtained with this type of IRM (von Davier, Gonzalez & Mislevy, 2009; Mislevy *et al.*, 1992b). Unless a correct characteristic of a population or subpopulation was incorporated into the population model using conditioning variables, it has been shown that plausible values will generally not recover the distribution correctly, and will produce biased results if used for the secondary analysis. Mislevy *et al.* (1992b) stated that “*Simple regression analyses based on nonincluded variables, for example, exhibited an average of 15% shrinkage*” (p. 155). Whereas if the imputation model is correctly specified using conditioning variables, plausible values provide users with data sets for secondary analysis to obtain unbiased estimates of parameters (Mislevy, 1984; Wu, 2005).

Rubin (1987) described the analysis with multiple imputations (plausible values in our case) as a three-step process. The first step involves creating M multiple imputations for each of the missing observations. In this case, this means using the Monte-Carlo method to approximate the integrations by randomly drawing M plausible values from Equation 2.20 for each student. Each of the M sets of plausible values from all students becomes a “completed” dataset that would be used as an intermediate step for estimating population parameters or for secondary analysis. The second step involves analysis of each of the M datasets in turn, to estimate population parameters. This will provide M sets of parameter estimates. Finally, the M sets of parameter estimates are combined according to the following formula to be the final estimate of population parameters:

$$\hat{t} = \frac{1}{M} \sum_{m=1}^M \hat{t}_m \quad \text{Equation 2.25}$$

Where M is the number of plausible values and $\hat{t}_m$ is a population parameter estimate from the m^{th} plausible values (imputation) dataset.

Each of the plausible values provides unbiased estimates of population parameters. However, it is important that the population parameter t must be estimated first with each of the plausible values as if the plausible values were the true values of θ , then take the average of the M estimated parameters to be the final estimation $\hat{t}_m$ of the population parameter t (Rubin, 1987; von Davier *et al.*, 2009). If M plausible values from each individual student are averaged first, this would result in a point estimate for individual student that is similar to EAP, and will result biased population variance and percentile estimates (von Davier *et al.*, 2009).

To estimate the standard error of a parameter estimate, one needs to take into account its sampling variance and the between imputations variance (measurement variance). The sampling variance ($\bar{U}$) is due to population parameters that were estimated from sampled students. This can be estimated with jackknife (Efron and Stein, 1981), bootstrap (Efron, 1979) or other techniques for each set of plausible values and then averaged according to Equation 2.26. The estimation of sampling variance for each parameter estimate of this dissertation is given in Section 3.5.1. Sampling variance is estimated for each set of plausible values and then the average sampling variance ($\bar{U}$) is calculated as below:

$$\bar{U} = \frac{1}{M} \sum_m U_m \quad \text{Equation 2.26}$$

where U_m is the sampling variance estimate from the m^{th} imputation dataset

Student ability (θ) is an unobserved latent variable. Student ability distributions (the posterior distribution) are estimated using item parameters, student responses to the test administered and available student background information. The posterior distribution

represents the uncertainty involved in estimating student ability due to its latent nature. The randomly drawn plausible values (multiple imputations) from the posterior distribution $f(\theta_n; x_n, Y_n, \beta, \sigma^2)$ of student n represents the ability distribution. Each randomly drawn plausible value would likely result in different values. This implies there is a “sampling variance” (B), that is between imputation variance due to the randomness with the procedure. Equation 2.27 is used for computing the variance among the M sets of estimates to approximate the uncertainty due to the unobserved θ .

$$B = \frac{1}{M-1} \sum_{m=1}^M (\hat{t}_m - \hat{t})^2 \quad \text{Equation 2.27}$$

These two variances are then summed according to Equation 2.28 to obtain the final estimate of the standard variance (V) for $\hat{t}$, and its standard error is the square root of the variance (Equation 2.29). The factor $(1+1/M)$ in Equation 2.28 is an adjustment for finite M (Rubin, 1987; Schafer & Olsen, 1998). This method of estimated total variance for a population parameter estimate has been adopted in most of the large-scale assessments or surveys where plausible value methodology is used (Rubin, 1987).

$$V_M = \bar{U} + \left(1 + \frac{1}{M}\right) B \quad \text{Equation 2.28}$$

Standard error (SE) of a parameter estimate is:

$$SE = \sqrt{V_M} \quad \text{Equation 2.29}$$

2.4 The Initial Development of the Plausible Values Technnique

The plausible value technique was initially developed for reporting the results of the NAEP 1983-84 assessment (Mislevy *et al.* 1992a). NAEP is one of the largest assessment surveys of student achievement in America. The objective of the NAEP survey is to consistently estimate population characteristics. The assessment is a sample-based survey to approximate the corresponding population parameters and uses a Balanced

Incomplete Block-spiralled sampling design (Mislevy *et al.*, 1992b). With this type of sampling design, different assessment items are assigned to different students, and students are assigned different numbers of assessment items (between 2 and 40 in NAEP). This means some sampled students only complete a small proportion of items from the entire test item pool.

With this type of assessment design that is used for most large-scale assessment surveys, it is unlikely to produce a consistent population parameter estimates based on aggregating and analysing individual estimates using the first type of IRM as expressed in Equation 2.1. Mislevy *et al.* (1992a) has demonstrated that the traditional method using a two-step process to estimate population parameters will result in an over-estimated population variance even with efficient individual parameter estimates. In order to obtain unbiased population parameter estimates, the second type of IRM model specified in Equation 2.4 was utilised. With the second type of IRM model, the population parameters are directly estimated without the need to estimate individual parameters first. This procedure overcomes the shortcomings that are involved when aggregating individual student parameter estimates to approximate their underlying population estimates. To assist secondary analysis, multiple plausible values for each student are drawn and made available to researchers and analysts. After NAEP developed the technique, plausible values have become widely used in large-scale assessment surveys.

The plausible value method provides three main advantages when used for estimating population characteristics. First, plausible values give consistent estimates of population characteristics. Second, plausible values provide an estimate of the precision of measurement that takes into account both the sampling and the latent nature of variables. And third, with multiple imputations for every student, the plausible value method produces a complete database that is easily used for secondary analysis (Beaton, 1987; Mislevy *et al.* 1992a).

The use of plausible values has some disadvantages: more work is required to produce multiple sets of plausible values, more computer storage is required to store multiple sets

of data and more work is required to analyse multiple datasets. These disadvantages are inconsequential in present days with readily available computing power and efficient statistical software.

It is important that care is taken in specifying the population model (Equation 2.8) used for drawing plausible values. Plausible values must not be used to make a conclusion about a subgroup without being conditioned when drawing the plausible values (Rubin, 1987; Mislevy, 1991). It is also important that the appropriate procedure outlined by Rubin (1987) and described in Section 2.3 is followed when using plausible values for secondary analyses.

2.5 Where Have Plausible Values Been Used?

The plausible value technique developed and used for 1983-84 NAEP has been used in subsequent NAEP surveys and adopted by many other large-scale educational assessments to provide population parameter estimates. For example, the plausible value methodology has been used in the Trends in International Mathematics and Science Study (TIMSS) (Martin & Kelly, 1996), International Adult Literacy Study (Murray, Kirsch & Jenkins, 1998), Progress in International Reading Literacy Study (PIRLS) (Martin *et al.* 2003), Programme for International Student Assessment (PISA) (Adams & Wu, 2002) and more recently in the Australian National Assessment Program – Literacy and Numeracy (NAPLAN) (ACARA, 2015b). In the following sections, examples of some of the large surveys and assessments that have used plausible value methodology are described.

National Assessment of Educational Progress (NAEP)

NAEP is the largest nationally representative and continuing assessment of what American students know and can do in various subject areas, including mathematics, reading, science, writing, the arts, civics, economics, geography and U.S. history. NAEP

results are based on a representative sample of students who are administered a relatively small number of test items. In the 1983-1984 NAEP reading assessment, 2-parameter or 3-parameter IRM models were first fitted to response data and then a population-structure model (multiple regression model with conditioning variables) was used to construct five sets of plausible values. Achievement results are reported in terms of means, variance and proportion of students above a specific proficiency level for populations of students (e.g., all fourth-graders) and groups within those populations (e.g., female students, Hispanic students). NAEP does not provide achievement scores for individual students or results for schools. NAEP can report results for selected large urban districts. Five plausible values were used in the early years of NAEP. However, in the 1998 NAEP technical report, it stated that “for stability, the estimates of the between-imputation variance ... are based on 100 plausible values” (Zhang, Isham & Worthington, 2001, p. 304). According to the NAEP Technical Documentation Website (NAEP National Assessment of Educational Progress, 2016), the number of plausible values for each of the sampled students has been increased from 5 to 20, however it is not clear when this change occurred. The reported plausible values are linearly transformed from logit values to scale scores in the range between 0 and 500 (Beaton, 1987).

Trends in International Mathematics and Science Study (TIMSS)

TIMSS is an educational research study on student achievement in mathematics and science around the world. It is designed to measure and interpret differences in national educational systems in order to help improve the teaching and learning of mathematics and science worldwide. From 1994, TIMSS began the first study in a regular cycle of studies at four-year intervals. TIMSS assesses achievement in mathematics and science at middle primary and lower secondary levels and collects background information on student, classroom and school contexts through questionnaires. The TIMSS achievement test design also makes use of matrix-sampling techniques to divide the assessment item pool so that each sample of students responds to just a portion of the items, thereby achieving wide coverage of the mathematics and science subject areas while keeping the response burden on individual students to a minimum.

ACER ConQuest was used for the 1995 TIMSS scaling (Adams *et al.* 1997b). ACER ConQuest uses a multidimensional random coefficients logit model with the addition of a multivariate linear model imposed on the population distribution. From 1999, one, two or 3-parameter IRT models, depending on the item type and scoring procedure, were used for TIMSS scaling (Yamamoto, 2000). Plausible value methodology was used to achieve reliable indices of student proficiency distributions and for use in statistical analyses and to provide good estimates of parameters of student populations. Five plausible values were drawn for each sampled student in each cycle of TIMSS. The reporting plausible values are in scale scores with a mean of 500 and standard deviation of 100 in the first cycle in setting the scale.

Progress in International Reading Literacy Study (PIRLS)

The Progress in International Reading Literacy Study (PIRLS) is an international study of reading achievement. PIRLS and TIMSS are both conducted by the International Association for the Evaluation of Educational Achievement (IEA). PIRLS started in 2001, and was designed to measure children's reading literacy achievement, to provide a baseline for future studies of trends in achievement and to gather information about children's home and school experiences while they are learning to read. PIRLS adapted the same methodology (from the various IRT models) used in TIMSS for scaling and drawing of plausible values for the sampled student proficiency estimation. Similar to other studies, five plausible values per sampled student were drawn for each cycle of PIRLS. The reported plausible values were transformed to scale scores with a mean of 500 and a standard deviation of 100 overall across all participating countries in the first cycle in setting the scale.

The Programme for the International Assessment of Adult Competencies (PIAAC)

The International Adult Literacy Survey (IALS, 1994-1998) and The Adult Literacy and Lifeskills Survey (ALL, 2003 and 2008) recently have become the Programme for the International Assessment of Adult Competencies (PIAAC, 2012). All of these are

cyclical, large-scale studies of adult skills and life experiences focusing on education and employment. A variation of matrix sampling techniques were used in these assessments. The scaling models used in these assessments were the 2-parameter logistic model (Birnbaum, 1968) for dichotomously scored responses and the generalized partial credit model (GPCM) (Mislevy *et al.* 1992b) for items with more than two response categories. Interestingly, ten plausible values were drawn for each respondent in PIAAC. There was no reason given in the PIAAC technical report (OECD, 2013) regarding the use of ten plausible values instead of five plausible values as in other similar large-scale assessments, e.g. IALS and ALL, etc. The reported plausible values were transformed on to the reporting scale with the mean between 237 and 277 and the standard deviation around 52 depending on the reporting scales (OECD, 2013).

Programme for International Student Assessment (PISA)

The Programme for International Student Assessment (which started in 2000) was designed to assist governments to monitor the outcomes of education systems in terms of student achievement on a regular basis (every three years) and within an internationally accepted common framework. In this way, PISA helps governments to not only understand, but also to enhance the effectiveness of their educational systems and to learn from the educational practices of other countries. PISA collects information on student achievement and contextual information about students, teachers and schools in a three-year cycle. The mixed coefficients multinomial logit model as described by Adams, Wilson and Wang (1997c) was used to scale the PISA cognitive data (up to PISA 2012) using an earlier version ACER ConQuest software (Wu, Adams, Wilson & Haldane, 2007) developed by the Australian Council for Educational Research. Up until the 2012 cycle, five plausible values for each sampled student were drawn from the population model and used for subsequent secondary analyses. For PISA 2015, however, the scaling was carried out with both 1-parameter and 2-parameter IRM models, and the number of plausible values drawn increased to ten per sampled students (OECD, 2017). Detailed reasons for this increase are not reported in the PISA 2015 Technical Report. von Davier (private communication, 2016) indicated that the reason for using a larger number of

plausible values was to provide better estimates of the variability when a large amount of imputation is required. The plausible values were transformed to scale scores for reporting.

National Assessment Program – Literacy and Numeracy (NAPLAN)

The National Assessment Program – Literacy and Numeracy (NAPLAN) assessments have been undertaken by students in Years 3, 5, 7 and 9 across Australia since 2008. NAPLAN results are reported using five national achievement scales, one for each of the four assessed aspects of literacy – Reading, Writing, Spelling, Grammar and Punctuation and one for Numeracy. The NAPLAN achievement scales were constructed with the IRT model using the ACER ConQuest software (Adams *et al.* 2015). Each NAPLAN scale spans Years 3, 5, 7 and 9, thus making it possible to monitor progress of students over time. NAPLAN provides two types of proficiency estimations on each assessment scale: maximum likelihood, using Warm's (1985) Weighted Likelihood Estimator, and five plausible values for each student. The Weighted Likelihood Estimates (WLE) were used for reporting student and school level performance. The plausible values were used to estimate population and subpopulation parameters for reporting and comparing student' performance at the national, state and subgroup level. The reported plausible values were linearly transformed to the scale scores with a mean of 500 and standard deviation of 100 across all four year levels in 2008 to set each domain reporting scale.

In summary, all the large-scale assessments mentioned above were first scaled with IRT models, from a 1-parameter logistic model up to a 3-parameter IRT model, in order to establish the assessment scale. The plausible values were drawn by fixing on the estimated item parameters and conditioning on the population model. The population parameters were estimated with the plausible values. The multiple imputation procedure described by Rubin (1987) was subsequently applied to the plausible values for estimating population parameters. With all these large-scale assessments, the plausible values have been linearly transformed to scale scores by multiplying a large factor. Table 2.1 provides a summary of the use of plausible values in these large-scale assessments described above.

Table 2.1 Examples of Plausible Values Used in Large-scale assessments

Assessment Program	Domain Tested	Target Population	Scaling Model	No. of PVs	Stats Reported
NAEP (US National testing Program)	mathematics, reading and science etc.	4 th , 8 th and 11 th grades	One, two or three parameter IRM	5 before 2016, 20 in 2016	means, variances and proportions of students above specified proficiency levels
TIMSS (Trends in International Mathematics and Science Study) (PIRLS)	Mathematics and Science	4 th and 8 th grades	One, two or three parameter IRM	5	Mean, Standard deviation, percentiles,
Progress in International Reading Literacy Study	Reading literacy	9-year-olds (4 th grade) students	One, two or three parameter IRM	5	Averages and percentages of student against international benchmarks
PIAAC (The Programme for the International Assessment of Adult Competencies)	Literacy, Numeracy and Problem solving	16 to 65 year-old adults	two parameter IRM	10	Average scores and percentage by achievement levels
PISA (OECD testing program)	Mathematics, Reading and Science	15 years old students	Generalised Multidimensional IRM up to 2012. One and two parameter IRM for 2015 & 2018.	5 before 2015, 10 in 2015	Mean, variance and proportion of students in specified bands for OECD, country and subgroups
NAPLAN (Australia National testing program)	Reading, Writing, Spelling, Grammar and Punctuation, and Numeracy	students in Years 3, 5, 7 and 9	Generalised Multidimensional IRM	5	Mean, standard deviation, percentage of students in specified bands for National, state and subgroups

2.6 How Many Plausible Values Should Be Used for Secondary Analysis?

Plausible values can be seen as multiple imputations for missing data when student latent ability is considered to be missing data. They are random draws from the marginal posterior distribution (Equation 2.12) for each student. These random draws are used to approximate the integration in the denominator of Equation 2.12, using the Monte-Carlo integration. When plausible values are used in the Monte-Carlo estimation, there is no precise answer for how many of them should be used. Typically, estimation software (for example, ConQuest 4 (Adams *et al.* 2015)) would draw 2,000 plausible values or more for approximating population characteristics. With regards to secondary analysis, as in plausible values technique first developed and used for NAEP (Beaton, 1987; Mislevy *et*

al. 1992a), most large-scale assessments adopted the same practice of providing five plausible values (Martin & Kelly, 1996; Martin *et al.* 2003; Adams & Wu, 2002; ACER, 2010) for secondary analyses. Only until recent years, there is a trend emerging in increasing the number of plausible values used for secondary analysis to more than five (NAEP National Assessment of Educational Progress, 2016; OECD, 2017).

This dissertation is concerned with the number of plausible values used in secondary analysis of large-scale assessment results. As described in the previous section, plausible values are multiple imputations for the missing latent proficiency of each student in the IRM context, when the student latent proficiencies are cast as missing data. There is very little research found that directly studied the number of plausible values that should be used for secondary analysis in the IRMs context. Two research papers found have a small section in studying the effect of number of plausible values used for secondary analysis. These were discussed later in this section.

There have been research studies on the number of multiple imputations investigating multiple imputations for missing data relating to regression, multiple regressions, structure equation modelling or various multilevel modelling work (Schafer, 1997; Collins, Schafer and Kam, 2001; Graham, Olchowski and Gilreath, 2007; Bodner, 2008; von Hippel, 2018). The missing data referred in these studies relate to missing data in respondents background information or measurement data of a multiple-cycles study that is missing for some participants in particular cycles. For example, information on parent education is missing for some students in a regression study when parent education is one of independent variables or predictors; or achievement data were missing for some absent students in a data collection cycle in a multi-year longitudinal study. It should be noted that there is a difference between missing observable data and student latent proficiency being considered missing in the IRM context. In the observable data case, multiple imputation for missing data is based on values of other independent variables or other participants with similar background information while in the IRM context, information about student latent proficiency (the missing data) is strongly related to the observed student responses to the instrument administered to the student. This implies that the

fraction of missing information in IRM context could be small. In the context of missing observable data, there have been several research studies regarding fraction of missing information (Rubin, 1987; Schafer, 1997; Longford, 2005).

Rubin (1987) derived formulae for the number of imputations required to meet set precision requirements. Rubin showed that together with the number of imputations M , the fraction of missing information, γ , influences the relative efficiency of parameter estimates. If the inference is a valid randomisation-based inference of the complete data and the imputation method is properly formed, the efficiency (relative to an estimate from an infinite number of imputations) of an estimate based on M imputations, is approximately $(1 + \gamma/M)^{-1/2}$ in standard error units. For example, with 90% missing information and 5 imputations, relative efficiency is approximately 92%. Rubin also stated that in cases with little missing information, proper imputation with $M = 2$ or 3 is nearly fully efficient (relative to an infinite number of imputations) and yields accurate coverage (the coverage is t -based interval estimates). However, when the fraction of missing information is very large, Schafer suggested that more than 10 imputations may be needed (Schafer, 1997). Rubin (1987) mentioned that in the simplest case, the fraction missing information could have the same value as the fraction of missing data. In multivariate data, however, the fraction of missing data may not equivalent to the fraction of missing information (Longford, 2005).

The Rubin's fraction of missing information is an index that gives the relative increase in estimate error variance due to nonresponse (Schafer & Olsen, 1998). In other words, the fraction of missing information is the increase in standard error caused by the missingness. Longford (2005) defined missing information as:

The missing information about a parameter is defined as the difference of the information contained in the complete and incomplete data sets, and the fraction of missing information is the ratio of this difference and the complete data information (p. 55).

The fraction of missing information is generally unknown. Rubin (1987) showed one way to estimate of the fraction of missing information (γ) for a parameter is:

$$\gamma = \frac{r + 2/(df + 3)}{r + 1} \quad \text{Equation 2.30}$$

$$\text{where} \quad r = \frac{(1 + M^{-1})B_M}{\bar{U}} \quad \text{Equation 2.31}$$

Pan and Wei (2016) conducted a research to demonstrate how the fraction of missing information (γ) determined with various fraction of missing data. The data used for Pan and Wei study was a 3-year cycle survey data from 2567 respondents. The fraction of missing data were at 4%, 10%, 20% and 29% in the responses. From this study, they showed that the fraction of missing data were strikingly small (less than 0.01%) with 99 imputations used regardless the size of fraction of missing data. From the research, they have concluded that the fraction of missing information and the fraction of missing data have different meaning and are not comparable.

Bodner (2008) demonstrated that precision (e.g. p values, confidence intervals and estimated fraction of missing information for an estimated parameter) improves by increasing the number of imputations in the analysis of incomplete data (again, missing observable data). The empirical data of Bodner's study analysed the three sets of cholesterol measurements from 28 patients (Bodner, 2008). Each cholesterol measure were obtained at a different time. The missing data occurred in the third set of measurements which 9 out 28 patients cholesterol measurement were missing. Bodner showed that the variability of estimated p values, confidence intervals and fraction of missing information were decreased with an increase of number of imputation. He argued that this would lead to categorically different inferential conclusions about estimation precision and null hypothesis rejection when use small number of imputations compared to large number of imputations. However, Bodner (2008) concluded that estimated parameters and estimated sampling variances of the parameters “*will exhibit less variability across imputations when few cases have missing values and when variable*

with missing data are strongly related to variables with observed data in the imputation model (p. 673)”.

There had not been research regarding fraction of missing information in the IRM context, where the fraction of missing data is 100%. Equation 2.30 shows that the missing information is related to degree of freedom (df). Applying this to the IRM context in large-scale assessments where the degree of freedom is generally large, Equation 2.30 becomes:

$$\gamma = \frac{r + 2/(df + 3)}{r + 1} \approx \frac{r}{r + 1} \quad \text{Equation 2.32}$$

when M is large:

$$r \approx \frac{B_M}{\bar{U}}$$

therefore, the Equation 2.32 becomes: $\gamma \approx \frac{B_M}{B_M + \bar{U}}$ Equation 2.33

In the context IRMs, reliability (ρ) is an index often used to describe the properties of an instrument. Reliability is often defined as the proportion of observed variation in ability estimates that is true (see Adams, 2005). Adams showed that in the context of marginal item response models, the uncertainty increase in the population parameter estimates is a function of a test reliability index. In fact, the reliability is the expression of the same notion as the Rubin's γ . This means that the more reliable the test, the smaller would be the fraction of missing information, and hence the fewer would be the number (M) of plausible values required for estimating population parameters.

There is relatively little guidance found in the literature as to the number of plausible values to be used with respect to population parameter estimate based on the IRMs. A recently published paper by Luo and Dimitrov (2019) compared the individual point EAP estimated using Markov Chain Monte Carlo (MCMC) within the structural equation modelling framework with the EAP estimated using the MMLE. The paper concerns the number of plausible values needed for estimating individual EAP estimates using the

MCMC method. The focus of this paper (Luo & Dimitrov, 2019) is different from that of this dissertation. This dissertation is concerned with estimating population parameters and how many plausible values are required for that purpose, particularly in the context of reporting system results from large scale assessments.

Monseur and Adams (2009) conducted a simulation study which explored some limitations of using plausible values to report the results for a large-scale assessment, such as PISA. Although their focus was not on the number of plausible values, they showed various results using one plausible value and five plausible values among other proficiency estimates, e.g. MLE, WLE or EAP. There was no difference between the estimated results from the two sets of plausible values where the same model was used to generate plausible values (with or without conditioning). They also indicated that the bias appeared to be proportional to the number of items in test, which indicates bias is influenced by the reliability of the test.

A small simulation study by Wu (2005) showed that the plausible values recover the generating values well, in terms of the population mean and variance, even with one set of plausible values. Wu conducted the simulations over three conditions: a 20-item test taken by 2,000 students with on-target condition (matching the average of test difficulty to the average of student abilities), a 20-item test taken by 2,000 students with off-target condition (mismatching the average of test difficulty to the average of student abilities), and a 3-item on-target test with 2,000 students. Wu also examined the recovery of population percentiles in a 20-item on-target test. The results from Wu's simulations showed that five plausible values recovered the generating values well compared to using WLE, MLE or EAP. Often, even one plausible value per student was sufficient to recover population parameters (mean, variance and percentiles) adequately. The simulations, however, were conducted only under three conditions. In practice, one plausible value would not be able to provide the estimation of uncertainty due to missing information.

In most large-scale assessments, five plausible values have been generated for each student, for estimating population parameters and for constructing the database for

secondary analysis. In recent years there has been a tendency to increase the number of plausible values to more than five, as for NAEP 2016 and PISA 2015, with little explanation provided.

In conclusion, different numbers of plausible values have been used across various large-scale assessments, from 5 to up to 20. No published systematic research study has been found on the number of plausible values that should be used in large-scale assessments.

2.7 Research Questions

The plausible value technique was developed to estimate population parameters for large-scale assessments. As presented in previous sections, the use of plausible values is the implementation of a Monte-Carlo integration approach to the estimation of population parameters, where the plausible values are random draws from student's posterior distribution $h(\theta_n)$. Many factors could affect the population parameter estimates using the plausible value method. The following equation illustrates estimating the population mean (μ) as an example:

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \bar{\theta}_n \quad \text{Equation 2.34}$$

Combined with Equation 2.17 and the Monte-Carlo integration approximation,

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \bar{\theta}_n = \frac{1}{N} \sum_{n=1}^N \left[\int \theta_n h(\theta_n) d\theta_n \right] \approx \frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{m=1}^M \theta_{nm} \quad \text{Equation 2.35}$$

where M is number of random draws, e.g. the number of plausible values, and N is sample size, θ_{nm} is the m^{th} plausible value for student n .

rearranged Equation 2.35 becomes:

$$\hat{\mu} = \frac{1}{M} \sum_{m=1}^M \frac{1}{N} \sum_{n=1}^N \theta_{nm} \quad \text{Equation 2.36}$$

In Equation 2.36, term $\left(\frac{1}{N} \sum_{n=1}^N \theta_{nm} \right)$ is the estimated population mean from the m^{th} set of plausible values. It reveals that the number of plausible values M , the sample size N , and the posterior distribution $h = (\theta_n)$, which is related to the assumed prior distribution and the IRM model, play a role in the estimation and may have an effect on estimating the population parameters. This leads to a number of research questions.

Equation 2.36 shows that the population parameter estimate directly involves the number of plausible values. Until now, according to the literature review, most large-scale surveys using IRM applications, with a few exceptions, have used five plausible values for estimating and reporting population characteristics. One small-scale research study (Wu, 2005) has demonstrated that in the IRM model applications it is possible to use as few as one plausible value to report the population mean and variance with adequate accuracy. Of course, more than one plausible value is required to evaluate variation of the estimates due to unobserved latent nature since estimating standard error requires more than one estimates.

This leads to the first research question: *How would varying the number of plausible values used affect the population parameter estimates?* This question is directly related to considering the important question of *how many plausible values used are adequate for estimating population parameters?* For example, is it possible to estimate population parameters with fewer than five plausible values without sacrificing the precision of the estimated parameters? Should different numbers of plausible values be used for estimating different parameters, for example, five plausible values for estimating population parameters and 100 plausible values for estimating the uncertainty, as in NAEP 1998?

Since population parameters in large-scale assessments are often estimated from a sample, this will incur sampling error. In addition, measurement error must be evaluated when a survey involves specific measurement models. The total error consists of sampling and measurement error and provides guidance of the accuracy on the parameter estimates. It has been communicated that the reason for increasing the number of plausible values from 5 to 10 for PISA 2015 is to improve the precision of the standard error of estimates (von Davier, private communication, 2016). Hence the second research question relates to standard error of parameter estimates: *Will the size of the standard error of parameter estimates be reduced by increasing the number of plausible values used?*

One of the assumptions for valid inference using multiple imputations is that the imputation model is properly formed, that is, in the IRT context, the measurement model is based on the IRM and population models. It is known that the shape of the likelihood of $h(\theta_n)$ in Equation 2.35 would have an effect on the precision of the likelihood estimates. The shape of likelihood relates to the test length. In general, increasing test length would make a likelihood shape narrower and more peaked. A narrow and peaked likelihood will produce a small within variance, the variance among plausible values. This leads to the third research question: *How does the IRM reliability affect the population parameter estimates when estimated using plausible values, in terms of test length (number of items) and test targeting conditions?*

Plausible values are random draws from the individual posterior distribution $h(\theta_n)$. The posterior distribution $h(\theta_n)$ is proportional to the product of the likelihood and an assumed prior distribution $f_{\theta}(\theta|\alpha)$. The fourth research question relates to prior distribution assumption and explores factors that may affect the estimates of population parameters when a prior distribution assumption is violated using plausible values. That is: *How does an assumption of prior distribution affect the parameter estimates using plausible values?* What are the implications of a violation of the assumptions concerning the prior distribution? For example, does the precision of estimated parameters reduce with skewed distributions, especially for the parameters at the tail of the distribution, e.g.

95th percentile? Will an increase in the number of plausible values have any effect on the precision of the estimated parameters?

Sample size (N) in Equation 2.36 also plays a role in estimating population parameters. This leads to the question: *How does sample size affect the population parameter estimates using a different number of plausible values?*

To answer these research questions, simulation studies have been conducted under different conditions with data generated from the models with and without using student background variables as conditioning variables. In addition, studies with the empirical data from NAPLAN 2015 Numeracy tests were carried out.

The objective of this research is to evaluate the current common practice of using five or more plausible values in large-scale assessments to infer population parameters and to estimate the standard error of the parameters. This dissertation aims to fill some of the gaps in the research with respect to the number of plausible values to be used for secondary analysis in the IRT context. The ACER ConQuest 4 software (Adams *et al.* 2015) was used for these studies, including data generation for the simulations, item calibration and generating plausible values. ConQuest is a computer program developed at the Australian Council for Educational Research. ConQuest is available to and has been used by the wider measurement and research community to fit the most up-to-date item response models and for drawing plausible values. Large-scale assessments, such as PISA (before the 2015 cycle), TIMSS, PIRLS and NAPLAN, etc. have all used ConQuest to fit item response models and drawing plausible values. The ConQuest syntaxes used in the dissertation are available in the Appendix 2 for anyone interested in checking/duplicating this dissertation work. The data analyses in this dissertation study were carried out using SAS®/STAT software (SAS Institute Inc., 2015). SAS is one of most popular statistical analysis softwares. It has been widely used around the world across vast number of fields. It is possible to use other software, including those implemented in open source software, such as R to fit item response models and to conduct statistical analysis.

Chapter 3 Method

Two simulation studies and two empirical studies were undertaken for this dissertation. Simulation One and Simulation Two aimed to explore a range of simulation conditions that could affect the population parameter estimates as discussed in Chapter 2. In a typical simulation study, a series of pre-generated datasets with known parameters are analysed under a host of different statistical conditions. The purpose of a simulation is to determine how a particular statistical model behaves under various known conditions and whether its properties operate in a way that is appropriate for particular research applications (Koehler, Brown & Haneusea, 2009). The empirical data studies is to look at the implications when using plausible values to estimate population parameters with real data.

There are five sections in this chapter. Section 3.1 provides a description of the IRM model that was used for data generation and for drawing plausible values. Section 3.2 includes a description of the simulation study design and rationale. Section 3.3 describes the data generation procedures used for simulation studies are explained. Section 3.4 provides a description of how the simulation results will be evaluated and Section 3.5 describes the method used in the empirical data study.

3.1 The Model Used in This Dissertation

In this dissertation, all generated student response data for the simulation studies and the empirical data study with the NAPLAN data were dichotomously scored. Hence, the item response model used for this study is a simple logistic model. The model is implemented through ACER ConQuest 4 (Adams *et al.* 2015). The model is expressed as:

$$\Pr(X_{ni} = 1; \delta_i | \theta_n) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \quad \text{Equation 3.1}$$

where: $\Pr(X_{ni} = 1; \delta_i | \theta)$ is the probability of a student n response of item i ; θ_n is the ability of a student; δ_i is the parameter of item i .

The simulation data were generated assuming the simple logistic model with a fixed set of item parameters and predefined population parameters. The item parameter values are provided in Appendix 1. The generated response data were then fit to the simple logistic model coupled with a regression model using ACER ConQuest 4 (Adams *et al.* 2015), and the details of data generation is described in Section 3.3. As described in Chapter 2, there are two components in this model as shown in Equation 3.2: the first component is item response probability model that is the simple logistic model (Equation 3.1) and the second component is a population model.

$$f(x_n; \xi, \alpha) = \int_{\theta} f(x_n; \xi | \theta) f_{\theta}(\theta | \alpha) d\theta \quad \text{Equation 3.2}$$

The posterior distributions $h(\theta)$ of student abilities (defined as in Equation 2.12) were estimated and plausible values were drawn from the posterior distributions.

The general population model can be expressed as:

$$\theta_n = \sum_{j=0}^J Y_{nj}' \beta_j + E_n \quad \text{Equation 3.3}$$

where θ_n is the ability of student n , Y_n is a vector of y , values that are fixed and known values for student n , β_j is the corresponding vector of regression coefficients (the conditioning variables) and E_n is an error term that is independently and identically distributed with a mean of zero and a variance σ^2 .

When all $\beta_j = 0$, Equation 3.3 becomes:

$$\theta_n = E_n \quad \text{Equation 3.4}$$

Equation 3.4 is the formula used for the data generation without covariates in the first part of the simulation.

3.2 Simulation Design

The aim of the simulation studies in this dissertation are to investigate how many plausible values used are sufficient to recover the generating population parameters with adequate statistical properties. In order to assess how well the generating parameters are recovered under each of the simulation conditions; bias and root mean squared error (RMSE) of population parameter estimates, means, variances, percentiles and percentages within a proficiency band range were estimated, and standard error of each parameter estimate was also estimated.

The simulations were carried out in two parts. Simulation One was carried out with the data generated with the population model expressed by Equation 3.4, that is no covariates used and all population regression coefficients $\beta_j = 0$. Simulation Two was carried out with the data generated including covariates in the model with non-zero regression coefficients as expressed in Equation 3.3. Four factors: test length, test targeting, population distribution and sample size, were varied in the simulation studies. Table 3.1 shows the levels of the four factors in the simulation design and the details of data generation is given in Section 3.4.

Table 3.1 Design of the Simulation Studies

Factors	Simulation One		Simulation Two	
	Levels	Number of Levels	Levels	Number of Levels
Test Length	10; 20; 40; 60 & 80	5	10; 20; 40 & 60	4
Test Targeting	On-Target; Off-Target by 0.5; 1.0 & 2.5 logits	4	On-Target; Off-Target by 0.5 & 1.0 logits	4
Population Distribution	Normal; Uniform; Bimodal & Chi-squared ($df = 5$)	4	Normal	1
Sample Size	200; 2,000 & 10,000	3	2,000	1
Regression Coefficients	n/a		-0.09, -0.57 & -1.00	3

The aim of the Simulation One is to explore the effects of increasing the number of plausible values used with various conditions. The plausible values were drawn only using student response data without conditioning variables. The four factors manipulated at various levels in Simulation One are as follows:

1. *Test length* – five tests of lengths 10, 20, 40, 60 and 80 items were used to simulate a range of test lengths, from a very short test to a relatively long test. Large-scale assessments, such as the PISA and NAEP sample surveys, often use fewer than 20 items for reporting dimensions. In other settings, for example, cohort assessments such as NAPLAN uses as many as 64 items for reporting dimensions. This is why a range from 10 to 80 items was chosen. In general, test length is related to the test reliability (Lord, 1959). Increasing test length will generally improve the test reliability providing other test conditions are approximately the same. This aims to address research question three outlined in Chapter 2: whether a similar effect will be observed from using different numbers of plausible values with increasing the test lengths. In other words, to investigate if there is an interaction between test length and the number of plausible values used when estimating population parameters. It is expected that with an increase in test length, a test becomes more reliable, hence the measurement error of the test becomes smaller. It is possible that fewer number of plausible values used with a longer test could result in similar accuracy for parameter estimates as that from a shorter test.
2. *Test Targeting* – the targeting of a test is operationalised as the difference between the average of item difficulties and the average of student abilities. If one wants to optimise the estimates of population parameters, it is typically recommended that the test information function matches the shape of student ability distribution (Lord, 1959). However, in practice, test targeting can deviate from the ideal. For example, PISA is administered in a large number of countries, where the average of student proficiency levels can vary significantly across countries. A reasonably targeted assessment for a country or countries that have their average proficiencies close to the OECD mean, may well be too difficult for a country with an average proficiency that is well below the OECD

mean. Similarly, the NAPLAN tests have been administered in eight jurisdictions. On average, Northern Territory students have a much lower proficiency level compared to the students in other jurisdictions, hence the NAPLAN tests in general are very difficult for most of the students in the Northern Territory. In order to address research question 3 in Chapter 2: to study the effects of increasing the number of plausible values used for parameter estimations under various test targeting conditions, four test-targeting conditions were simulated in this study:

- tests with difficulty matching average student ability (on-target)
- tests off-target moderately by 0.5 logits
- tests off-target by 1.0 logit
- an extreme off-target case: tests off-target by 2.5 logits

Figure 3.1 demonstrates the four targeting conditions using a 40-item test with a normally distributed population. There is one graph for each of the four targeting conditions. On each graph, the population distribution is shown on the left, and the item distribution is shown on the right (each item shown as a triangle). With the extreme off-target condition, the simulations were only carried out for medium and large sample sizes. This is because item responses in a test that is extreme off-target by 2.5 logits over a small sample size ($N = 200$) are often all correct or all incorrect. Such items will provide no information in the Rasch modelling context and are excluded from the model estimation. Another aspect of the research question addressed here is: whether there is an interaction between the numbers of plausible values used for estimating population parameters and the test targeting conditions?

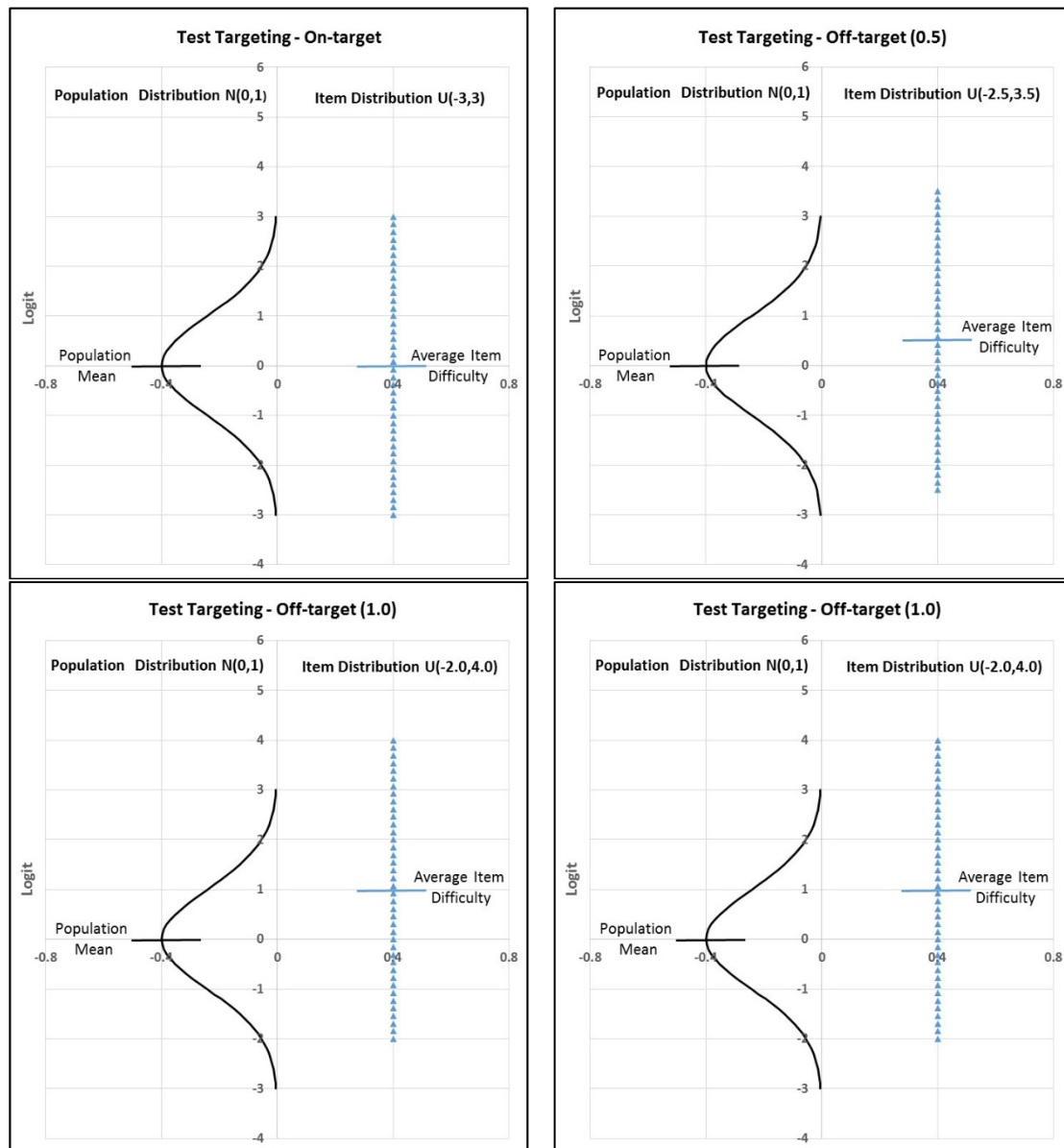


Figure 3.1 Test Target Conditions

3. *Population distributions* – A common criticism of marginal maximum likelihood estimation is that it requires additional assumption concerning the population distribution of latent variable. The bases of the concern is that an error in the assumption could have add adverse affect on the quality of the parameter estimates (Mislevy *et al*, 1992b). With marginal maximum likelihood estimation, although any population distributions can be applied, for reasons of simplicity, normally distributed or conditional normally distributed populations are often assumed (Mislevy *et al*, 1992b). However, for a real population, a

latent trait distribution can deviate from this normality assumption and have a skewed distribution. In addition, when there are two sub-populations combined, for example, Indigenous students and non-Indigenous students, this could result in a bimodal population distribution. In order to investigate the effects with different population distributions, four population distribution shapes (Figure 3.2) were chosen to generate the simulation data in this part of the study. They were a normal distribution, bimodal distribution, uniform distribution and chi-squared distribution with 5-degrees of freedom⁸ to simulate a skewed sample distribution. In order to facilitate the comparison among different simulation conditions, the four sample distributions were all set to have a mean of zero and standard deviation of one. Thus, the distributions are: normal distribution $N(0, 1)$, uniform distribution $U(-\sqrt{3}, +\sqrt{3})$, bimodal distribution $[N(-\sqrt{3}, 0.2) \text{ and } N(+\sqrt{3}, 0.2)]$ (a combination of two normal distributions with means of $-\sqrt{3}$ and $+\sqrt{3}$ respectively, and a standard deviation of 0.2) and chi-squared ($df = 5$) distribution which was a standardised chi-squared distribution with five degrees of freedom. For the chi-squared distribution, a vector of sample size N from a chi-squared ($df = 5$) was drawn first and then standardised. This resulted in a vector of random sample (N observations) with a mean of zero and variance of one. The chi-squared ($df = 5$) distribution was positively skewed (skewness of 1.26) which the other three distributions were symmetric (skewness of zero).

⁸ chi-squared distribution with 5-degrees of freedom is expressed as chi-squared(5) in Figures and expressed as chi-squared ($df = 5$) in the text or tables of this dissertation.

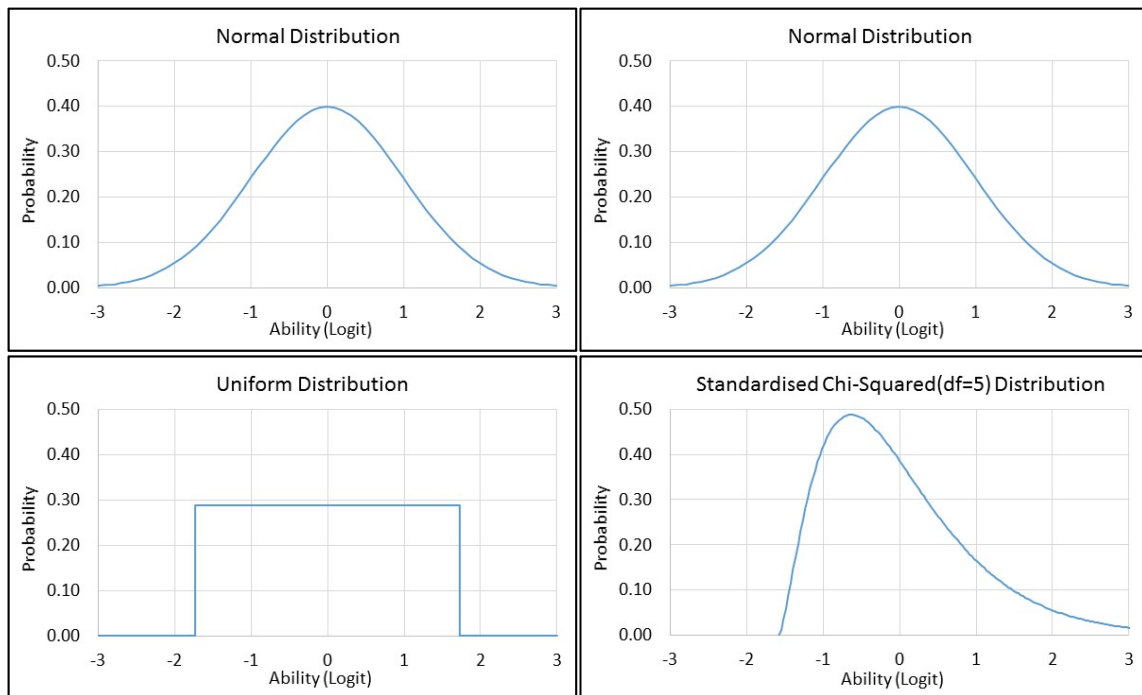


Figure 3.2 Sample Distributions Used for Simulation One Data Generation

4. *Sample size ($N = 200, 2000$ & $10,000$)* – three sample sizes were used to simulate small, medium and large samples. A small sample size of 200 cases reflects the situation when there is a small subgroup or a limitation with the sampling process, this could be because of sampling financial restrictions, timing issues or any other limitations. For example, trial tests are typically conducted with relatively small sample sizes. With an increase in sample size, the sampling error reduces, hence the variability of sample statistics will be reduced. A medium sample size of 2,000 is a reasonable sample size for a simulation study as it is large enough to provide reliable estimates and the simulations can be conducted within a reasonable time frame. A large sample size of 10,000 was used to reflect the situation when sampling error is not a major concern. Further increasing the sample size will significantly increase simulation time. By varying sample size the research questions that are aimed to be answered are: How does sample size affect the recovery of population parameters? Will the accuracy of population parameter estimates increase with the same number of plausible values used when sample size increases?

The aim of Simulation Two is to evaluate if there is any effect of increasing the number of plausible values used to recover the population parameters and regression coefficients, and also to compare with the findings of Simulation One. The main difference between Simulation One and Simulation Two is the assumption in data generation. With Simulation One, it was assumed that sample data were from a homogeneous population, no covariate was used for the data generation. Simulation Two assumed that the sample data were from a heterogeneous population. Three regression coefficients (β_j in Equation 3.3) were specified in Simulation Two for these covariates. These three regression coefficient values were chosen to closely resemble the regression coefficients of background variables based on analyses of the NAPLAN 2015 student data (ACARA, 2015b). For the data generation, three independent variables (Female, LBOTE and Indigenous) were included as covariates in the model.

$$\theta_n = \beta_0 + \beta_1 * Female + \beta_2 * LBOTE + \beta_3 * Indigenous + E_n \quad \text{Equation 3.5}$$

Regression coefficients were given values as follows:

$$\beta_0 = 0$$

$$\beta_1 = -0.09$$

$$\beta_2 = -0.57$$

$$\beta_3 = -1.00$$

E_n is the error term, independently and identically normally distributed $N(0, \sigma^2)$.

A representative sample of the student background data were selected from the NAPLAN 2015 the Northern Territory data and used for Simulation Two. The sample was selected with specific characteristics of a small sample with a high proportion of Indigenous students. The gender, LBOTE and Indigenous status in the selected data have no specified correlations. In the data, female and male students have an approximately equal proportions, and about 30% are LBOTE students and about 30% are Indigenous students. Test length and test targeting condition were manipulated in this part of the simulation. Four test lengths (10, 20, 40 and 60 items) and three test targeting conditions (on-target, small off-target and moderately off-target by 1.0 logit) were used. All simulations in this

part were carried out with a medium size sample ($N = 2,000$) that was sampled from a normally distributed population as shown in Figure 3.3.

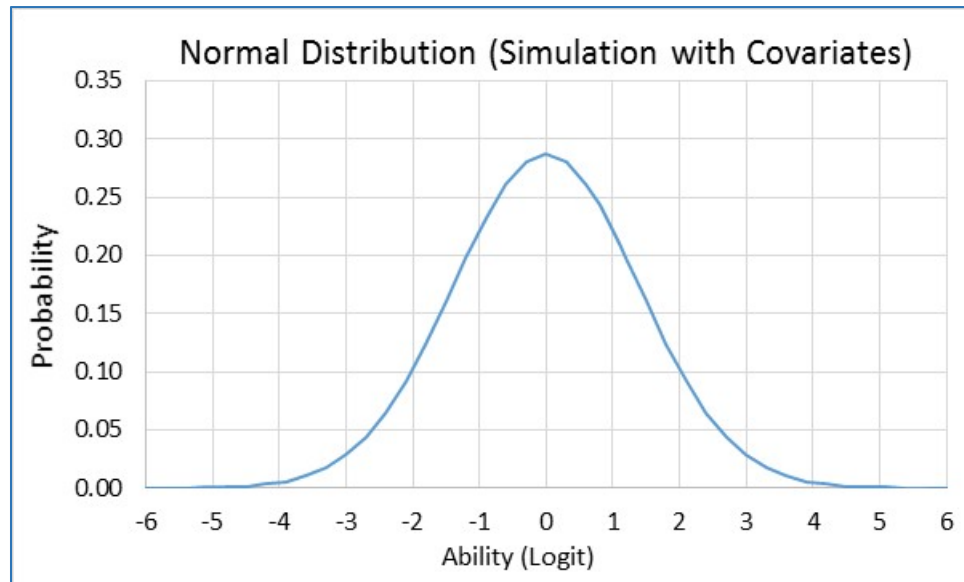


Figure 3.3 Sample Distribution Used for Simulation Two Data Generation

Overall, Simulation One yielded a $4 \times 4 \times 4 \times 3$ matrix, that is, test length by test targeting by population distribution by sample size, which results in a total of 220^9 combinations of conditions (cells). Simulation Two yielded a $4 \times 3 \times 1 \times 1$ matrix, that is, four different test lengths as those in Simulation One by three test targeting status (on-target, off-target by 0.5 and off-target by 1.0 logit) by one population distribution (normal distribution) by one sample size ($N = 200$), a total of 12 combinations of conditions (cells).

Since the item parameter estimates are not of interest in this study, item parameters have been kept as fixed. Item parameters are uniformly distributed for all simulations within the range of 6 logits for all tests (Appendix 1).

⁹ For the off-target test by 2.5 logits, the simulation is only conducted for medium and large sample sizes.

3.3 Identify the Number of Replications Used for Simulations

Monte-Carlo simulation is a numerical technique of experimenting with random sampling from probability distributions. In a typical simulation study, a series of computer-generated datasets are analysed under a host of different statistical conditions. The purpose of simulation is generally to determine how a particular statistical model behaves under various conditions, and whether its properties operate in a way that is appropriate for particular research applications. In simulation studies, the datasets are computer-generated based on the population generating model specified by the researcher. Hence, the generated sampling distribution is known for the particular set of conditions. This allows researchers to test how good the model reflects the generated data (Castro & Gaviria, 2000; Axelrod, 2003) and to investigate a wide variety of “what if” questions.

As a simulation involves repeated experiments based on generated data from a pre-defined model, it is expected that a slightly different result will be obtained from each replication of a simulation (Sokolowski & Banks, 2009). Increasing the number of replications generally results in smaller errors of parameter estimates. However, a large number of replications in simulation studies can be time consuming and expensive. Therefore, it is important to assess how many replications are required to obtain a specified level of accuracy in the results when conducting a simulation study.

There are two sampling components in the data generation of each replicate in the simulation studies of this dissertation. First, student abilities were drawn from an assumed population distribution, and second, item responses of a student were drawn based the simple logistic model with a set of fixed item parameters. For each replication, one set of sample statistics can be obtained. Theoretically, the number of replications in a simulation can be increased to an infinite number. In practice, however, simulations can only be conducted for a limited number of replications. The number of replications depends on available resources. With an insufficient number of replications, a simulation may produce estimates of population parameters with a large error which could lead to incorrect conclusions (van Belle, 2008). In order to be able to summarise the result across

replications with a degree of precision and accuracy that suits the purpose of a study, the number of replications used for the simulation study needs to be determined.

The intention of the simulation studies in this dissertation is to be able to detect a bias or RMSE that is equivalent to the size of one tenth of a sampling error, $|Bias|$ or $|RMSE| \geq \frac{SE}{10}$, therefore, the number of replications required can be determined as follows:

For the estimated bias $\hat{b}$ in the Mean estimate over G replications:

$$\hat{b} = \frac{\sum b_g}{G} \quad \text{Equation 3.6}$$

where: b_g is bias in an estimated mean from replication g .

The variance over replications is:

$$Var(\hat{b}) = Var\left(\frac{\sum b_g}{G}\right) = \frac{\sigma^2}{NG\rho} \quad \text{Equation 3.7}$$

where: $Var(\hat{b})$ is the variance of the estimated bias, N is the sample size, ρ is the reliability (the EAP/PVs reliability) of the test.

When $|\hat{b}| \geq 1.96\sqrt{\frac{\sigma^2}{NG\rho}}$, the bias is significant at 0.05 level.

In order to detect a bias or RMSE of mean estimates that is equal to or greater than one tenth of the sampling error (SE):

$$\frac{SE}{10} = \frac{1}{10} \frac{\sigma}{\sqrt{N}} \geq 1.96\sqrt{\frac{\sigma^2}{NG\rho}} \quad \text{Equation 3.8}$$

$$G \geq \frac{19.6^2}{\rho} = \frac{384}{\rho}$$

Hence G needs to be at least $384/\rho$. From the ACER ConQuest 4 (Adams *et al.* 2015) simulation outputs the (EAP/PVs) reliability are between 0.489 and 0.920. The variation in the reliability is mostly related to the test length. The longer the test is, the higher will be the reliability of the test. This gives the minimum number of replications as 786. It was decided that 1,000 replications be conducted for each condition (cell) of the simulation studies.

3.4 Data Generation

The steps involved in data generation for both simulation studies are as follows:

1. *Specify item parameters*

A set of item parameters were specified and remained fixed during data generation for tests with the same number of items. Six item parameter files were created, one for each test length. Each file contained a different number of items (See Appendix 1 Item Parameters for Simulation Studies). The item parameter values were equally spaced between -3 and $+3$ logits. For the simulations with each off-target condition, the difference between average item difficulty and average student ability was added to each set of the parameters.

2. *Specify population distribution parameters*

For Simulation One, the population parameters for each distribution were specified for generating student data. For Simulation Two, apart from the population parameters, the regression coefficients (as β_1 , β_2 and β_3 in Equation 3.3) were also specified, and a 2000-by-3 covariate matrix was created. This matrix has 2,000 cases each with three dummy coded indicators, one for each independent variable, as each independent variable has only two categories. For easy calculation and interpretation of the outcomes, each covariate variable was coded to have the sum equals to zero across 2,000 cases. The residual error

variance was independently and identically distributed as $N(0, 1)$, thus the mean from the generated sample equaled zero. The population variance was then the sum of the variance of three covariates and the variance of the regression residuals which is one.

3. *Draw random samples*

A random sample of θ vectors of appropriate size was drawn from an assumed population according to Equation 3.3, in which E_n was independently and identically distributed as $N(0, 1)$. The algorithm used for uniform deviates (uniform random number generator) in ACER ConQuest 4 was described in Chapter 7 of *Numerical Recipes In C* (Press, Teukolsky, Vetterling & Flannery, 1992) and the algorithm used for normal deviates was described by Box and Muller (1958).

4. *Combine random latent vectors and predefined item parameters*

Random latent vectors θ and predefined item parameter values ξ were combined to calculate the probability of each item and each simulated student. These probabilities were then compared to a random number generated from uniform $(0, 1)$ to allocate responses to a specific item according to the rule:

$$X_{ni} = \begin{cases} 1 & \text{if } P(\theta_n) \geq \text{random number from uniform } (0, 1) \\ 0 & \text{other wise} \end{cases}$$

Steps 3 and 4 were carried out with the "Generate" command in ACER ConQuest 4. There were no missing responses in the generated response data for the two simulation studies.

5. *Fit the generated response data to the simple logistic model*

Fifty-four plausible values per generated student were drawn with the population model using ACER ConQuest 4. Examples of the command file for data generation, item calibration and drawing plausible values are given in Appendix 2.

6. *Replications*

Steps 3 to step 5 were replicated 1,000 times for each simulation design cell (see Table 3.1). A date time parameter was used as a seed number for the random number generators for each data generation replication.

3.5 Evaluation of the Simulation Outcomes

The evaluation of the simulation outcomes was carried out in two main stages: estimation of the population parameters of interest and evaluation of how well these population parameters were recovered by using indices of bias and RMSE.

3.5.1 Estimate Population Parameters

For each replication, 54 plausible values per simulated student were drawn that formed 54 datasets. The 54 sets of plausible values were divided into six groups. The groups consisted of 1, 3, 5, 10, 15 and 20 numbers of plausible values. A parameter of interest (t) was first estimated separately using each of the 54 plausible values and then the results were aggregated for each of the six groups following the method described by Rubin (1987) as show in Equation 3.9:

$$\hat{t} = \frac{1}{M} \sum_{m=1}^M \hat{t}_m \quad \text{Equation 3.9}$$

Where M is the number of plausible values, $\hat{t}_m$ is the estimate of the parameter of interest based on the m^{th} set of plausible values, $\hat{t}_m$ is the final estimate of t .

Each of the six groups of results were aggregated with the respective number of plausible values across a whole sample as below:

Group 1: consisted of one set of plausible value, e.g. 1st plausible value

- Group 2: consisted of three sets of plausible values, e.g. the 2nd to 4th plausible values
- Group 3: consisted of five sets of plausible values, e.g. the 5th to 9th plausible values
- Group 4: consisted of ten sets of plausible values, e.g. the 10th to 19th plausible values
- Group 5: consisted of fifteen sets of plausible values, e.g. the 20th to 34th plausible values
- Group 6: consisted of twenty sets of plausible values, e.g. the 35th to 54th plausible values

For most of the large-scale assessments, a range of population parameters, like means, variances and percentile ranges together with standard errors for the parameter estimates, are often estimated and reported. The global and regional assessments typically divide students into a number of proficiency levels and report the percentage of students against these levels (ACARA, 2015a). For these reasons, several population parameters, mean, variance, 90th, 95th and 97.5th percentiles, were estimated for each simulated condition. In addition, percentages within proficiency bands were also calculated for two bands divided by a cut score. The cut score was arbitrarily placed to be at the 80th percentile for each of the four generating population distributions in both simulation studies. The band for each of the 54 plausible values was determined. When a plausible value was equal to or above the cut score, it was deemed to be in the top band, while if a plausible value was below the cut score, it was deemed to be in the low performance band. Thereafter, the percentages in each of the two bands were determined. Because the percentages of students in these two bands always summed to 100 percent, only the percentage in the top band was reported.

Table 3.2 shows the generating parameter values and the derived parameter values for Simulation One with an on-target test across four sample distributions. These parameter values were the same across different test lengths. For different test targeting conditions, a shift that was equal to the off-target size was added to the item parameter values.

Table 3.2 Generating and Derived Parameter Values for Simulation One

Distribution	Generating Values (logit)		Derived Values (logit)			
	Mean	Variance	The cut score for Percentage in a band (at 80 th percentile)	90 th Percentile	95 th Percentile	97.5 th Percentile
Normal	0	1	0.8416	1.2816	1.6449	1.9600
Uniform	0	1	1.0392	1.3856	1.5588	1.6454
Bimodal	0	1	0.9407	1.0583	1.1463	1.2190
Chi-squared ($df = 5$)	0	1	0.7239	1.3397	1.9197	2.4769

Table 3.3 lists the generating parameter values and the derived parameter values for Simulation Two. The derived parameter values were computed based on the generating normal distribution. These parameter values were the same across different test lengths, again, a shift which equals the off-target size was added to the item parameter values for the off-target simulations.

Table 3.3 Generating and Derived Parameter Values for Simulation Two

Generating Values		Derived Values		
Mean	Variance	90 th Percentile	97.5 th Percentile	Cut Score ($\geq 80\%$)
0	2.3896	1.9788	3.0275	1.2988*

* It is used to determine the percentage at the top band (above the cut score)

Standard error of a parameter estimate: for each of the estimated parameters, its standard error was also estimated. In this context, there are two contributors to the overall error variance. One is the variance from student sampling which relates to the uncertainty of parameters estimated from samples rather than the underlying population. The other component is the measurement error variance due to the fact that student ability is a latent variable that is not directly observable and is derived from the student responses using an IRM. The variation among plausible values reflects the measurement error variance.

The total variance (V_M) of a parameter estimate is the sum of these two variance components. Rubin (1987) has provided a method of estimating the overall standard variance as shown in Equation 3.10:

$$V_M = \bar{U} + (1 + M^{-1})B_M \quad \text{Equation 3.10}$$

where: M is the number of plausible values used in estimation.

$\bar{U}$ is the average sampling variance over M sets of plausible values:

$$\bar{U} = \sum_{1}^M U_M / M \quad \text{Equation 3.11}$$

B_M is the variance among the M estimates:

$$B_M = \frac{1}{M-1} \sum_m (\hat{t}_m - \hat{t})^2 \quad \text{Equation 3.12}$$

The standard error of a parameter estimate is the square root of the total variance. This method of combining sampling variance and measurement variance has been adapted for most of the large-scale assessments that use plausible values to estimate population parameters (Bearton, 1987; Adams & Wu, 2002; ACARA, 2015b). With the simulation, the total variance of a parameter estimate can also be estimated from the variance of parameter estimates over replications which contains both the student random sampling variance and the variance related to latent ability estimates. For the studies in this dissertation, the Rubin (1987) method was used for estimating the total variance for each replicate, and the final total variance is the average of the total variance over the simulation replicates. The reported standard error is the square root of the final total variance.

The sampling error variance was estimated based on a simple random sample assumption for the simulation studies. The sampling error variance of mean was estimated as:

$$SamplingErrorVariance_{Mean} = \frac{\overline{S^2}}{N} \quad \text{Equation 3.13}$$

where $\overline{S^2}$ is the average of the estimated sampling variance for a particular number of plausible values, e.g. three plausible values, five plausible values, etc. and N is the sample size.

The sampling error variance of the variance was estimated as (Ahn and Fessler 2003):

$$SamplingErrorVariance_{Variance} = S^4 \left(\frac{2}{N-1} \right) \quad \text{Equation 3.14}$$

The sampling error variance for a percentile was estimated as below (Brown and Wolfe 1983):

$$SamplingErrorVariance_{Percentile} = \frac{1}{f^2(x_p)} \left(\frac{p(1-p)}{N} \right) \quad \text{Equation 3.15}$$

where: x_p is the p^{th} percentile, $f(x_p)$ is the probability density function at the p^{th} percentile and N is the sample size.

The sampling error variance for the percentage in a band was estimated through the sampling error variance for a proportion as below:

$$SamplingErrorVariance_{Proportion} = \left(\frac{p(1-p)}{N} \right) \quad \text{Equation 3.16}$$

where: p is the proportion concerned and N is the sample size.

The sampling error variance for the regression coefficients were estimated from the regression model using the procedure PROC REG in the SAS®/STAT software (SAS Institute Inc., 2015).

3.5.2 Evaluation of the Estimated Population Parameters

In order to evaluate the precision and accuracy of each population parameter's recovery, the bias and RMSEs of the parameter estimates, that result from using each of the six groups of plausible values, were calculated. The purpose was to examine whether the accuracy of the parameter recovery would improve if the number of plausible values used was increased. In addition, the sampling error variance and the measurement error variance was estimated for each set of plausible values (that is $M = 3, 5, 10$ or 20 ; No measurement error can be estimated for the first group which contains only one plausible value). The sampling variance and measurement variance were combined into a total variance for each of the estimated parameters. The following section describes how bias, RMSE and standard error of a specific parameter were calculated.

Bias is the difference between the generating value of a parameter and its estimate from the simulation, which is given by:

$$Bias(\hat{t}) = \left(\sum_{g=1}^{1000} (\hat{t}^g - t) \right) \div 1000 \quad \text{Equation 3.17}$$

where: $\hat{t}$ is an estimated parameter from the generated sample (mean, variance, percentiles, percentage in proficiency band and regression coefficients), and $\hat{t}^g$ is the estimation of the parameter from the g^{th} replication and t is the generating value.

The smaller the bias, the closer the parameter estimate is to its generating value and the more accurate the estimate.

RMSE stands for the root of the mean squared error (*MSE*), which is the variance of an estimated parameter around its generating values. This was estimated as the average squared difference between the generating value of a parameter and its estimation from the simulation. *MSE* consists of two components: squared bias and variance of the

parameter estimates, e.g. $MSE(\hat{t}) = Bias^2(\hat{t}) + Var(\hat{t})$, which indicates the overall accuracy.

RMSE is given by the formula:

$$RMSE(\hat{t}) = \sqrt{\sum_{g=1}^{1000} \frac{(\hat{t}^g - t)^2}{1000}} \quad \text{Equation 3.18}$$

The MSE values are always greater than zero. A value closest to zero means that there is greater accuracy in the estimates. For an unbiased estimated parameter, the MSE is the variance of the estimated parameter. In practice, RMSE is often reported instead of MSE because RMSE has the same unit as the parameter making it easier to interpret. The accuracy and precision of a parameter estimate related to its generating value is depicted in Figure 3.4. As shown in Figure 3.4, bias is the distance between estimated value and its generating value, it indicates how accurate the estimated value is. RMSE is the spread of estimated values which indicates the precision of the estimated value.

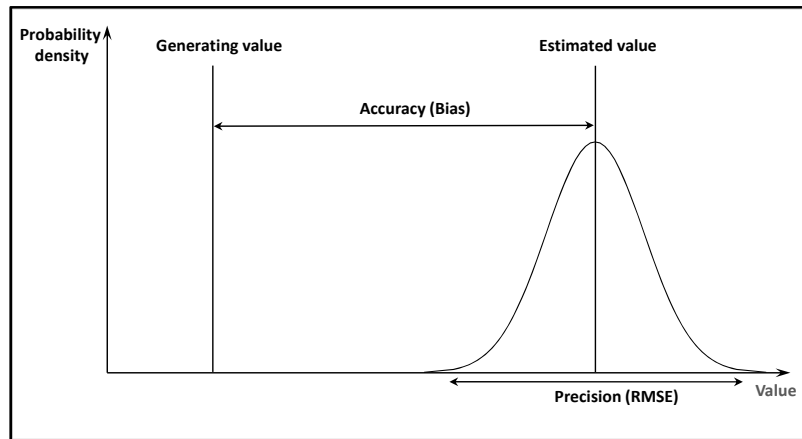


Figure 3.4 Illustration of Bias and RMSE

The following is a summary list of the steps for evaluating the simulation outcomes.

1. *Read plausible values*

Plausible values were read from each replicate for each of the simulated conditions. A total of 220,000 and 12,000 datasets were processed for Simulation One and Simulation Two, respectively.

2. *Estimate regression coefficients (Simulation Two only)*

For Simulation Two, to estimate regression coefficients β_j , each of the 54 vectors of plausible values were fitted to a multiple regression model specified by Equation 3.5 with the contrast variables. The contrast variables were defined in the data generation step 2. This was repeated for each of 54 sets of plausible values by replicates by simulation conditions. This resulted in a total of $54 \times 1000 \times 12$ regressions to be carried out. One set of regression coefficients was generated from each of 54 sets of plausible values.

3. *Estimate parameters*

Population parameters were estimated for each of 54 sets of plausible values by replicates and by simulation conditions. The parameters are mean, variance, percentage in the top band and percentiles.

4. *Aggregate parameter estimates*

The parameter estimates from 54 sets of plausible values were grouped into six groups, each group consisted of parameter estimates from a different number of plausible values: 1, 3, 5, 10, 15 or 20 plausible values.

5. *Estimate Bias and RMSE*

For each of the estimated parameters, its bias and RMSE were estimated for each of the 54 sets plausible values according to Equation 3.17 and Equation 3.18 and then combined into six groups results

6. *Estimate standard error*

Standard errors were estimated for the estimated parameters for five groups of results (excluding the group with one plausible value) by simulation conditions. The five groups of results were estimated using 3, 5, 10, 15 and 20 plausible values, respectively.

3.6 Illustration with the Empirical Data Studies

The second part of this study was carried out with empirical data. The empirical data used was the NAPLAN 2015 non-identifiable student response data. Permission was obtained from the Australian Curriculum Assessment and Reporting Authority for the use of the NAPLAN data for this part of the study. An ethics approval to conduct the research using the NAPLAN 2015 data for the empirical studies was obtained from the Human Research Ethics Committee of The University of Melbourne (Ethics ID Number: 1749625).

The NAPLAN tests are nationwide tests administered annually to all students in Australia in Year 3, 5, 7 and 9, which are over 1.1 million students each year. The NAPLAN tests assess five domains: Reading, Writing, Spelling, Grammar & Punctuation and Numeracy. Each year, the NAPLAN results are reported using five national achievement scales, one for each of the assessed domains. The NAPLAN student performance results include mean, standard deviation, percentiles and percentages in proficiency bands. Results are aggregated from five plausible values per student and reported at national and jurisdiction level, for the whole cohort and by subgroups, such as gender, Indigenous status and LBOTE (Language background other than English) (ACARA, 2015a). The plausible values are generated using the multidimensional linear regression model based on the response data and student background information. The details of constructing the NAPLAN scale and producing the plausible values can be found in the NAPLAN 2015 technical report (ACARA, 2015b).

The aim of this part of study is to investigate whether increasing the number of plausible values would have an effect on the reporting of results and whether the findings from the simulation studies will be reproduced with the empirical data. In order to be consistent with the simulation studies in this dissertation, the simple logistic model that is implemented with ACER ConQuest 4 was used to generate the plausible values from the NAPLAN 2015 data.

Instead of using data from all five domains, the empirical study was carried out with the NAPLAN 2015 Numeracy data. The reason for choosing Numeracy for this part of study is that the Numeracy tests across four year levels provide some variation in test length, while the tests in other domains have a similar number of items across the four year levels. For Numeracy, there are 35 items in the Year 3 test, 40 items in the Year 5 test, and 64 items in both Year 7 and Year 9 tests.

The analyses were carried out in two studies: Empirical Study One (whole cohort data) and Empirical Study Two (data of Indigenous students). For both studies, the analyses were carried out for the whole of Australia and by jurisdiction. There is a wide range of student abilities across the eight jurisdictions. The difference between the highest and the lowest average abilities in Numeracy is approximately 1.4 logits in Year 5, about 1.2 logits for Year 3 and Year 7, and about 1 logit in Year 9. The NAPLAN 2015 was a paper based test, every student in the same year level was administered the same test paper. This means that a test that was well targeted for one group of students could become a poorly targeted test for another group of students. Hence by examining parameter estimates from different groups, it may provide insight into the effect of test targeting conditions. Jurisdictional sample sizes in Empirical Study One were medium to large ranging between 3,023 and 307,549 students. The sample sizes in Empirical Study Two were much smaller ranging between 126 and 5,244. Hence, the analysis of Indigenous student data provides an insight into effects of small sample sizes.

The following describes the major steps that were used in this part of the study.

1. *Construct data files*

First, student response data files, one for each year level, were constructed to include regression contrast variables.

2. *Construct item parameter files*

Four Numeracy item parameter files were constructed, one for each year level. The item parameters used for drawing plausible values were obtained from the NAPLAN 2015 technical report (ACARA, 2015b).

3. *Draw plausible values*

The student response data files created in step 1 were fitted to the simple logistic model with item parameters fixed on those from step 2 using ACER ConQuest 4 and 100 plausible values per student were drawn.

4. *Estimate parameters of the NAPLAN 2015 cohort and subgroups*

Estimate the Numeracy statistics for the NAPLAN 2015 cohort and for Indigenous students, respectively. The statistics are those reported in the NAPLAN 2015 National Report, they are the mean, standard deviation, 20th and 80th percentiles, and the percentage at the top band by jurisdiction by year level for each of 100 sets of plausible values.

5. *Aggregate estimates with different number of plausible values*

The estimated parameters were combined into seven groups, each with a different number of plausible values. They are groups with 1, 3, 5, 10, 15, 20 and 46 plausible values, respectively.

6. *Estimate the standard error of each parameter estimate*

For each group which contains M sets of plausible values ($M = 3, 5, 10, 15, 20$ and 46), first, the sampling error was estimated for each set of plausible values. It was estimated by the PROC SURVEYMEANS procedure in the [SAS®/STAT] software (SAS Institute Inc., 2015) (ACARA 2015b) and averaged over M sets of estimated sampling errors to obtain the final sampling variance. Second, the measurement variance was estimated across M sets of plausible values. Then the sampling variance and the measurement variance were combined to obtain the total variance according to Equation 3.10, and the standard error is the square root of the total variance.

7. *Evaluate the effect of different number of plausible values used*

The parameter estimates based on using five plausible values were deemed as the reference values. The parameter estimates using M sets of plausible values ($M = 3, 10, 15, 20$ and 46) were then compared with the corresponding reference value.

3.7 Summary

This dissertation employed two simulation studies and two empirical data studies. Simulation One aimed to explore four factors that may affect the population parameter estimates using different number of plausible values. It was assumed that the data are sampled from a homogeneous population, and there was no covariate included in the model during data generation in Simulation One. Simulation Two aimed to explore the effect of increasing the number of plausible values used on population parameter estimations with sample data generated from a heterogeneous population. NAPLAN 2015 numeracy data were used for the empirical data studies, first with data from the whole NAPLAN cohort and then only using data from Indigenous students.

Both simulation data and empirical data were fit to the simple logistic model using ACER ConQuest 4: 54 plausible values per student (grouped into sets of 1, 3, 5, 10, 15 and 20 plausible values) were drawn for the simulation studies and 100 plausible values per student (grouped into sets of 1, 3, 5, 10, 15, 20 and 46 plausible values) were drawn for the empirical data studies. For the simulation studies, the bias and RMSEs of each parameter estimate using different numbers of plausible values were estimated and evaluated. For the empirical study, the results from five plausible values were used as reference values. Then the estimates based on different number of plausible values were compared with the reference value. The standard error for each parameter estimate were estimated for the simulation studies and empirical studies.

Chapter 4 Results of the Simulation Studies

The results from the two simulation studies are presented in this chapter. The difference between the two simulation studies is whether the underlying population is homogeneous or is heterogeneous for data generation, and Simulation Two was carried out with selected conditions that were used for Simulation One.

This chapter has two sections. Section 4.1 presents the outcomes from Simulation One. The bias and RMSE indices of parameter estimates, and the standard errors of each estimated parameter are presented. Six sets of results, each one aggregated with a different number of plausible values, were estimated for each parameter by simulation conditions.

Section 4.2 presents the outcomes from Simulation Two. Because the student background variables were used as conditioning variables while generating data in this study, the population parameters estimated included regression coefficients. Again the bias and RMSE indices, and standard errors of the estimated parameters were evaluated by simulation conditions and arranged into six sets of results. The bias and RMSEs of the regression coefficients are presented in Section 4.2.1. The bias and RMSEs of the other population parameter estimates are presented in Section 4.2.2. The estimated standard errors for the regression coefficients are presented in Section 4.2.3 and the estimated standard errors for the other parameter estimates (mean, variance, the percentage in the top band, 90th percentile and 97.5th percentile) are presented in Section 4.2.4.

4.1 Simulation One with Homogeneous Population

As described in Chapter 3, Simulation One was carried out by manipulating four factors during data generation. These four factors were test length, test targeting conditions, sample distribution and sample size. Several categories were varied for each of these four

variables: six test lengths, four test targeting conditions, four sample distributions and three sample sizes. This forms a total of 220 different simulation conditions¹⁰ (refer to Section 3.2). The steps involved in the simulation were:

- Generate data;
- Fit the data to the simple logistic model and draw plausible values;
- Estimate the population parameters;
- Evaluate how well the generating population parameter values were recovered from the simulation data by estimating the bias and RMSEs of parameter estimates; and
- Estimate the standard errors for each of the parameter estimates.

The simulation results were evaluated by estimating the population parameters that are listed in Table 3.2 for each simulated condition and then compared with the generating values. These parameters were mean, variance, percentage above the cut score, 90th, 95th and 97.5th percentile. The Bias and RMSE indices for each of the estimated parameters were then calculated and assessed. In addition, the sampling and the measurement error variances of each estimated parameter were also estimated and combined as a total standard error for the estimated parameters.

4.1.1 The Bias and RMSEs

For each of the simulation conditions, 1000 replications were conducted. For each of the 1000 replications, a total of 54 plausible values were randomly drawn for each generated student. After plausible values were drawn, each set of plausible values were analysed in the same manner for each of the six parameters of interest (that is: mean, variance, percentage in top band, 90th percentile, 95th percentile and 97.5th percentile). In order to

¹⁰ A sample size of 200 students with off-target by 2.5 logits are not simulated.

investigate the relationship between the number of plausible values used and the recovery of a generating parameter, the parameter estimates from each of 54 plausible values have been combined into six groups, each with a different number of plausible values.

The estimated parameters from each of the six groups were combined according to Equation 3.9 given by Rubin (1987). This resulted in six estimates $\hat{t}^g$ for the g^{th} replication for each parameter. This process was repeated for each of the 1,000 replications. The bias and RMSEs were then estimated according to Equation 3.17 and Equation 3.18, respectively, for each of the estimated parameters of interest by simulation condition (that is each condition in the simulation design matrix in Table 3.1).

The Bias

Bias is the difference between the expectation of a parameter estimate and the generating value. It was calculated as the average difference over 1,000 replications between the estimated value $\hat{t}$ and its generating value (t), according to Equation 3.17. The detailed bias in each parameter estimate are given in Appendix 3 and displayed in Figure 4.1.

Figure 4.1 contains six graphs, one for each of the six parameters. Each graph contains four panels, one for each of the four simulated population distributions. Within each panel, four different test targeting conditions are displayed: on-target, off-target by 0.5 logits, off-target by 1.0 logit and off-target by 2.5 logits. The three sample sizes used in each simulation are displayed in different colours: orange represents the smallest sample size of 200 students, blue represents the sample of 2,000 students and red represents the largest sample size of 10,000 students. Note that there is no orange dot for the extreme off-target condition (off-target by 2.5 logits) because this extreme off-target condition with small sample size ($N = 200$) was not simulated. The test lengths are represented by different symbols: a round dot (●) for the shortest 10-item tests, a diamond (◆) for 20-item tests, a triangle (▲) for 40-item tests, a square (■) for 60-item tests and a cross (+) for the longest simulated test, with 80-item. This colour and symbol scheme is used throughout all the graphs in this thesis.

Each plot in Figure 4.1 shows the bias in the parameter estimates calculated using M sets of plausible values ($M = 1, 3, 5, 10, 15$ or 20) by: test length, test targeting, sample distribution and sample size. The horizontal line at zero is a reference line. A marker above zero means that the bias estimated from M sets of plausible values was larger than its generating value; a marker below zero means that the bias estimated from M sets of plausible values was smaller than its generating value.

Figure 4.1 shows that when the number of plausible values used increased, the bias in each of the six parameter estimates showed very little difference within each of the simulation conditions. There was no increasing or decreasing trend observed with an increase in the number of plausible values.

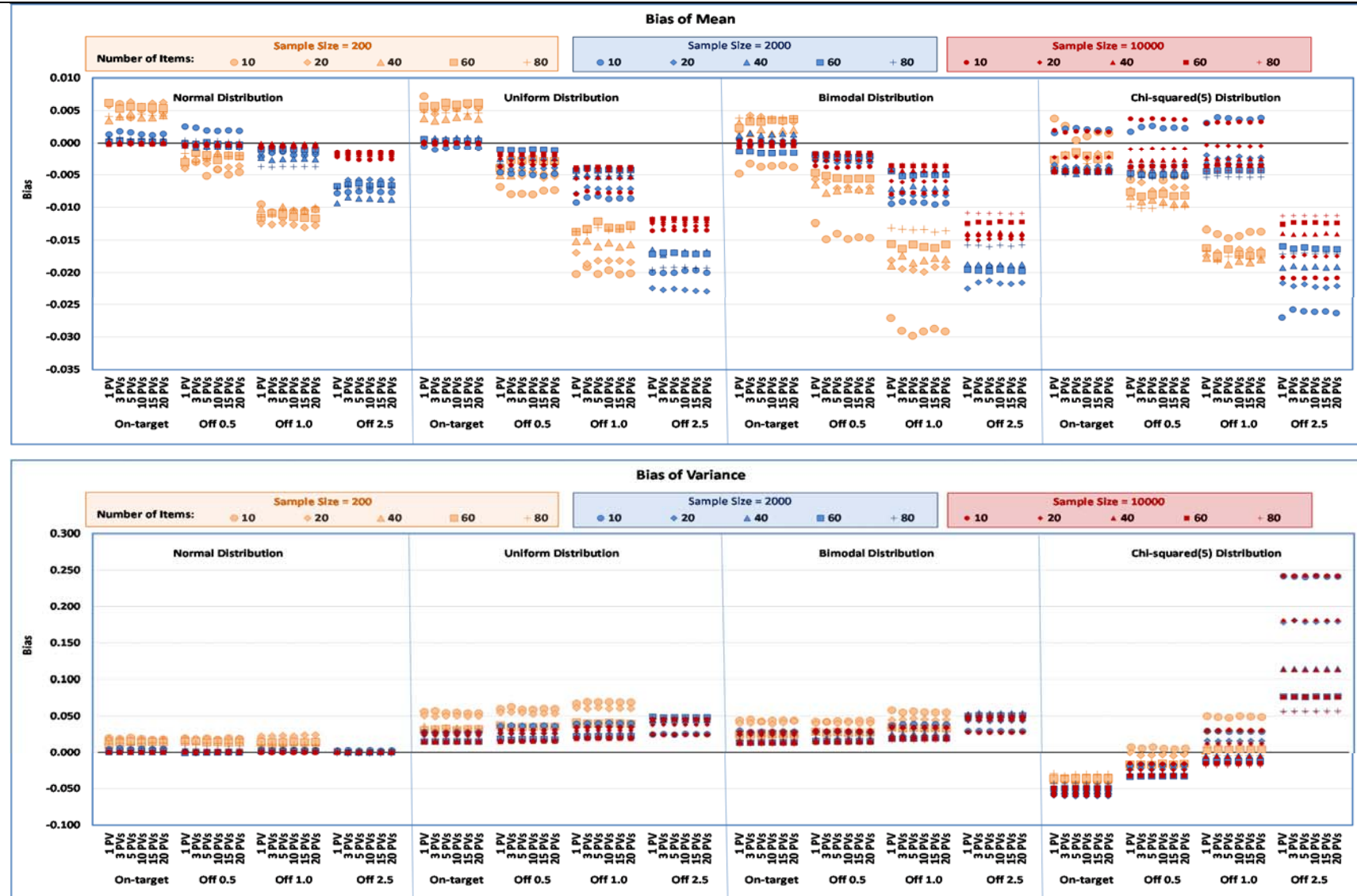
The effect of test length on the bias in the parameter estimates was not always consistent. Although increasing the test length tended to reduce the bias in the parameter estimates, there were exceptions, for example, a test with 80 items occasionally showed smaller biases than a 60 item test under the same simulation condition. The magnitudes of the bias in the three percentile estimates increased when the percentiles increased.

In terms of test targeting conditions, the results showed that the bias in the mean estimates increased with the increase in severity of the off-target condition, e.g. the more severe off-target test had a larger bias in the mean estimation for all simulation conditions. For the other five parameters, this trend did not always hold. For example, the bias in the three percentile estimates from the chi-squared ($df = 5$) distribution with the extreme off-target condition (2.5 logits) were smaller than that estimated with other better targeted conditions.

With respect to the sample distributions, the normally distributed samples had the smallest bias in all parameter estimates among the four different distributions when other conditions were kept the same. The skewed chi-squared ($df = 5$) distribution with extreme off-target condition had a much larger bias in the variance estimates compared to the bias in variance estimates for other distributions and targeting conditions. No clear patterns

were observed with the other two non-normally distributed samples. The bimodal distribution samples had the largest bias in the three percentiles estimates.

It was observed that the sizes of the bias in the estimated mean and variance were related to the sample sizes. That is the smaller the sample, the larger bias in the mean and variance estimates. The bias in the percentage estimates in the top band and the bias in the three percentile estimates were similar in magnitude across the three sample sizes, especially when the sample was from a normal distribution.





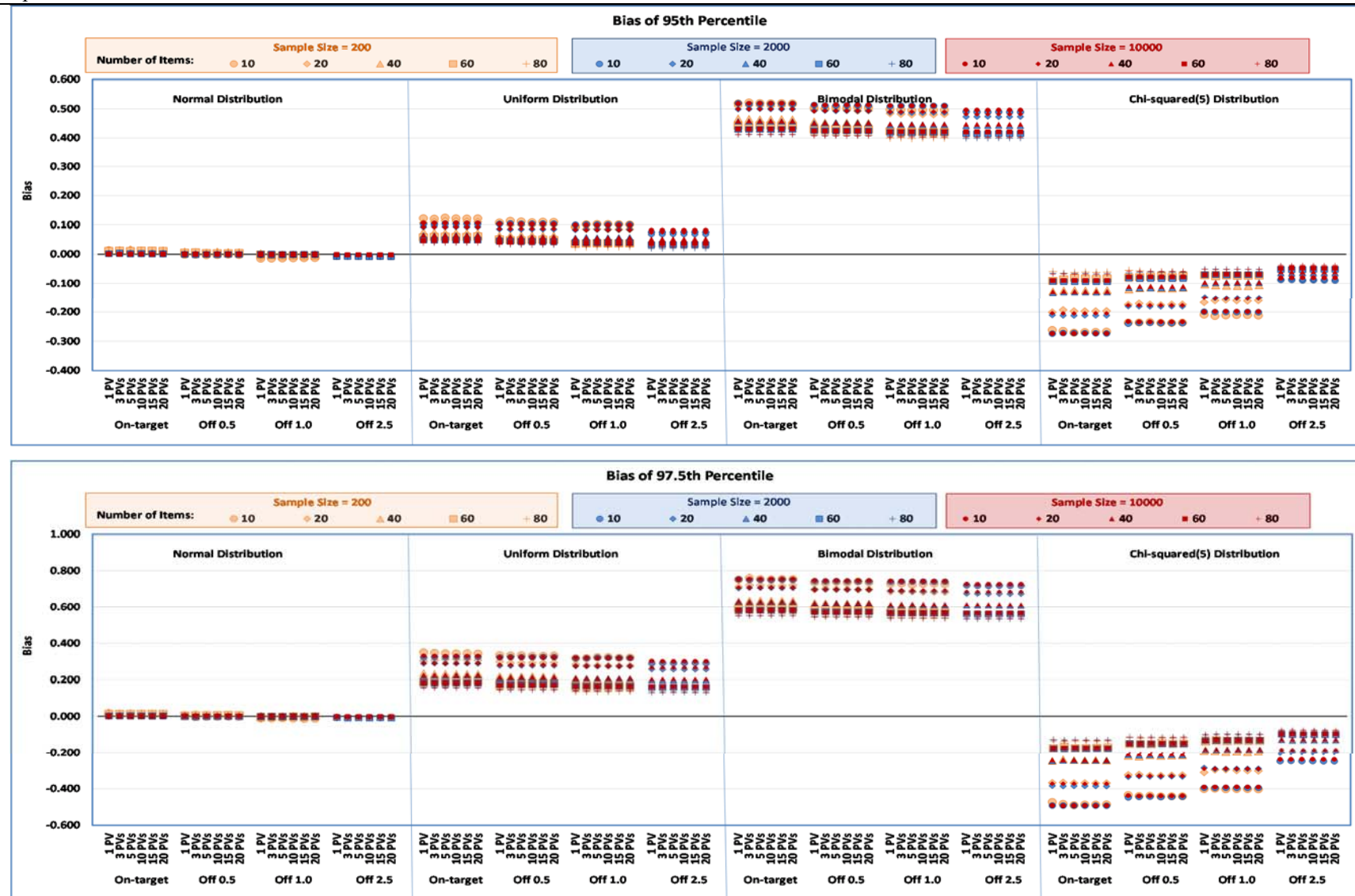


Figure 4.1 The Bias in the Parameter Estimates from Simulation One

The RMSE

RMSE quantifies the variability of the estimates around the generating value. The RMSE of a parameter was estimated according to Equation 3.18. The detailed RMSE results are presented in Figure 4.2 and also given in Appendix 3.

As in Figure 4.1, there are six plots in Figure 4.2, one for each parameter estimate. Each plot contains four panels, one for each sample distribution. Each plot shows the RMSEs that were calculated from six sets of plausible values by: test length, test targeting, sample distribution and sample size. The same colour and symbol scheme used in Figure 4.1 is also used for Figure 4.2. In order to make comparisons easier, the scale range between the maximum and the minimum are kept the same for the three percentiles plots.

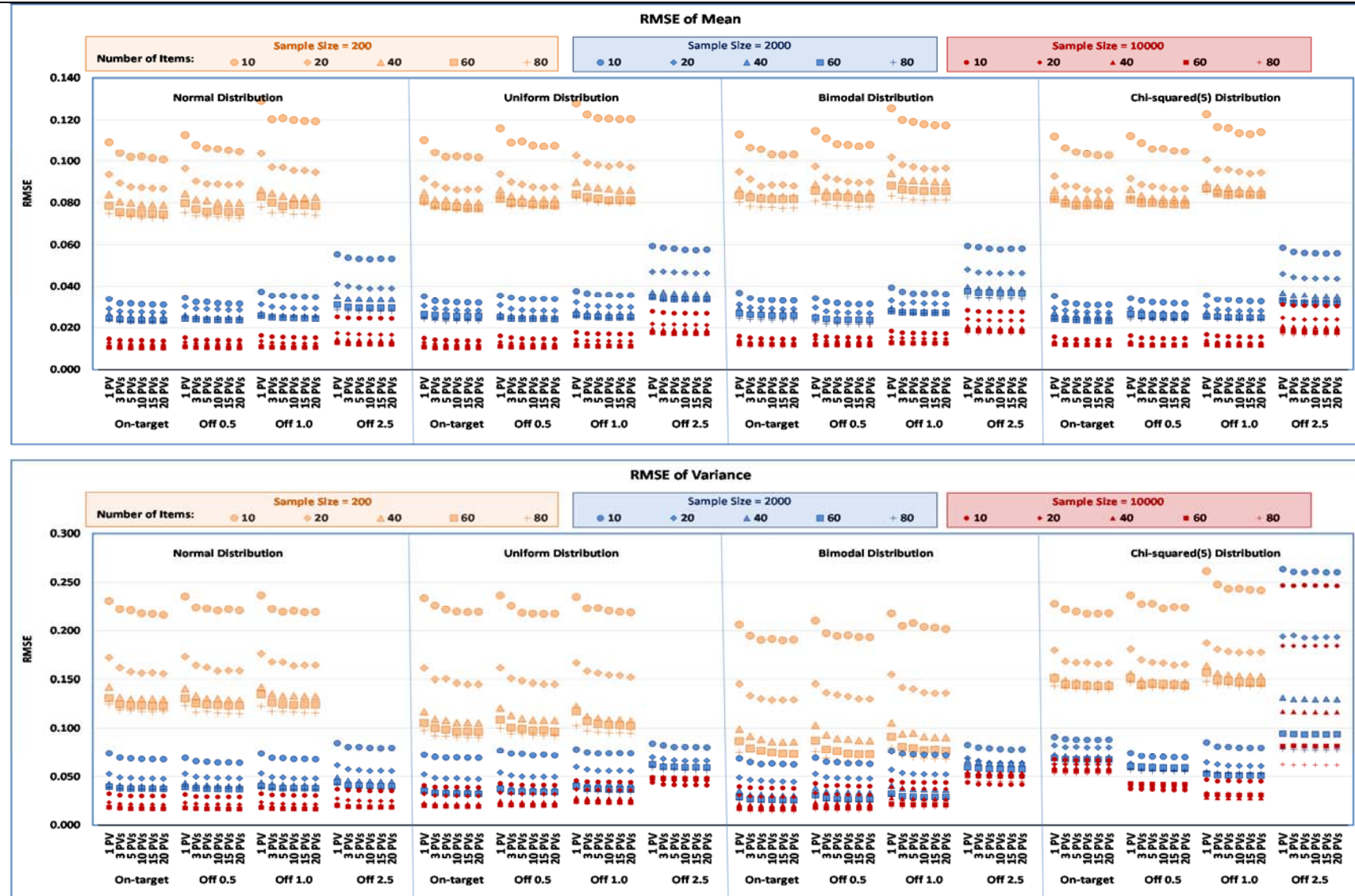
Figure 4.2 shows a similar pattern across all simulated conditions for all six parameter estimates. That is: the RMSEs of the parameter estimates reduced with an increase in the number of plausible values used. This pattern was the most obvious when the sample sizes were small ($n = 200$, orange coloured dots). The RMSEs were reduced with an increase in the sample size. And the magnitude of the reduction in the RMSEs with the number of plausible values decreased as the sample size increased. With a large sample size of 10,000, the differences among the RMSEs of estimates from one plausible value and estimates from 20 plausible values were smaller than 0.008 logits across all simulation conditions. With a small sample size of 200, the difference was up to 0.050 logits.

In relation to test length, it was observed that the RMSEs of the parameter estimates were larger with short tests and reduced when the test length increased. However, this effect was more obvious when the sample size was small and a test had fewer than 40 items.

In terms of test targeting status, when tests were reasonably targeted, that is the difference between average item difficulties and average student abilities was less than or equal to

1.0 logit, the RMSEs of the mean and variance estimates using the same number of plausible values were approximately the same when other simulation factors were the same. The RMSEs of the mean and variance estimates were larger when a test was extremely off-target compared with the RMSEs estimated from the other three targeting conditions. However, this pattern did not always hold for other parameter estimates. For example, for the sample from the chi-squared ($df = 5$) distribution, the RMSEs of the 95th and 97.5th percentiles that were estimated from the extreme off-target condition were smaller than the RMSEs of the 95th and 97.5th percentiles that were estimated from other three targeting conditions.

The population distribution had no effect on the RMSEs of the mean and variance estimates, but the RMSEs of the three percentile estimates from the bimodal distribution were much larger than the RMSEs estimated from other distributions.





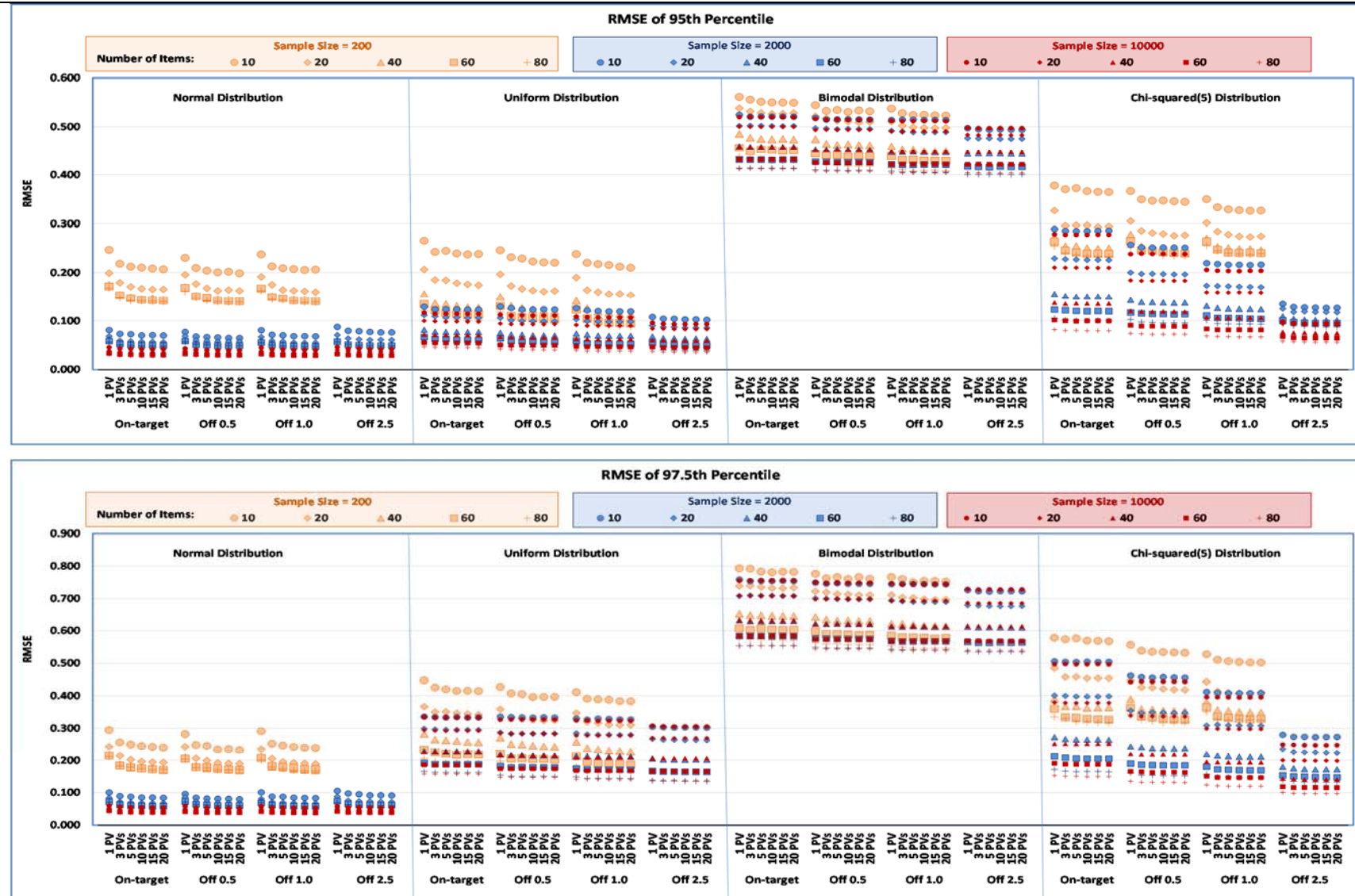


Figure 4.2 The RMSEs of the Parameter Estimates from Simulation One

4.1.2 The Estimated Standard Errors

There were two random sampling processes involved in these simulations, one was due to sampling of individual students and another was related to the student latent ability. The error variances from these two processes, the sample variance related to sampling students and the measurement variance related to latent ability estimates, were estimated as described in Section 3.5.1 and combined into the total variance for each parameter estimate using Rubin (1987) method. When sampling error variance and measurement error variance are estimated using plausible values by the Rubin (1987) method, the estimate of the standard errors is a function of the number of plausible values used (refer to Equation 3.10). This section examines the effect of the number of plausible values used on the standard error estimates.

The standard errors estimated in this dissertation were based on the method described by Rubin (1987) and discussed in Section 2.3 Working with Plausible Values. The steps involved in estimating standard errors for the estimated parameters are listed below:

- Estimate the sampling error variance of a parameter (Equation 3.13 to Equation 3.16) for each of 54 plausible values, then combine 54 sampling variance estimates into six groups according to Equation 3.11, each with a different number of plausible values;
- Estimate the measurement error variance using Equation 3.12 for each of the five groups of parameter estimates, each with a different number of plausible values;
- Combine the sampling and measurement variance estimates using Equation 3.10 to obtain the total variance, and the standard error is the square root of the total variance.

Table 4.1 shows the standard errors for the mean and variance estimates from a normally distributed sample of 200 students by the number of plausible values used; the tests are on-target with five test lengths ranging from 10 to 80 items. Table 4.1 illustrates the

effect of the number of plausible values used on estimating the standard error of population parameter estimates. It shows that both the standard errors of the mean and variance estimates reduced with increasing number of plausible values used. Also, the estimated standard errors reduced with increasing test length when other simulation factor were kept the same and the same number of plausible values was used. Table 4.1 shows that increasing the test length had a larger effect on the estimated standard errors than the effect from increasing the number of plausible values used. From Table 4.1, it can be seen that the estimated standard errors of the mean and variance estimates reduced by up to 3.6% and 5.9%, respectively, when increasing the number of plausible values used from 3 to 20. In contrast, the standard errors of the mean and variance estimates using three plausible values reduced 15.5% and 22.3%, respectively, when the test length increased from 10 to 80 items. The standard errors of the parameter estimates for each of the simulated conditions are provided in Appendix 4.

Table 4.1 The Estimated Standard Errors of a Normally Distributed Sample ($N = 200$) with Different Length On-Target Tests, by Number of Plausible Values

Standard Error	Test Length	3PVs	5PVs	10PVs	15PVs	20PVs	% Reduction**
Standard Error of Mean	10	0.0888	0.0879	0.0864	0.0859	0.0856	3.6
	20	0.0827	0.0820	0.0809	0.0806	0.0804	2.8
	40	0.0782	0.0773	0.0770	0.0768	0.0767	1.9
	60	0.0760	0.0757	0.0752	0.0751	0.0750	1.3
	80	0.0750	0.0744	0.0742	0.0742	0.0741	1.2
	% Reduction*	15.5	15.4	14.1	13.6	13.4	
Standard Error of Variance	10	0.1456	0.1407	0.1385	0.1368	0.1370	5.9
	20	0.1316	0.1292	0.1270	0.1266	0.1263	4.0
	40	0.1208	0.1192	0.1187	0.1176	0.1175	2.7
	60	0.1156	0.1142	0.1135	0.1134	0.1130	2.2
	80	0.1131	0.1121	0.1110	0.1108	0.1108	2.0
	% Reduction*	22.3	20.3	19.9	19.0	19.1	

* Percentage reduction in standard errors estimated between 10 and 80 items tests when the same number of plausible values were used.

** Percentage reduction in standard errors estimated from 3 to 20 plausible values used.

The patterns observed in Table 4.1 were also observed in other simulation conditions. Figure 4.3 shows the standard errors of estimated parameters by: sample size, test targeting and sample distribution. No standard error was reported for one plausible value. Again, there are six graphs in Figure 4.3, one for each of the six parameters. The colour

scheme and format in the Figure 4.3 graphs is the same as that used in the Figure 4.1 and Figure 4.2.

In recent years, it was suggested that five plausible values may not be enough for estimating standard errors (von Davier, Private communication, 2016), this argument has led to the trend of increasing the number of plausible values used in large-scale assessments (OECD, 2017). Figure 4.3 shows that the estimated standard error decreased with an increase in the number of plausible values used for all parameter estimates across all simulated conditions. The magnitude of the reduction was up to 5% between the standard errors estimated with three plausible values and with 20 plausible values. Across all simulated conditions there were a few exceptions, where there was a small increase in the estimated standard errors with an increase in the number of plausible values used. The exceptions occurred at the 95th and 97.5th percentiles from a chi-squared ($df = 5$) distribution small sample ($N = 200$) with a test having 60 or more items and extremely off-target status (off by 2.5 logits).

Figure 4.3 shows that the estimated standard errors were dependent on the sample size, that is the larger the sample size, the smaller the estimated standard error. With a large sample ($N = 10,000$), an increase in the number of plausible values used made very little difference on estimating standard errors regardless of the simulation conditions (the maximum difference in the estimated standard errors between using 3 and 20 plausible values were less than 0.0018).

Increasing the test length had a small effect when sample sizes were small or medium. In general the standard errors estimated for longer tests were smaller than the standard errors estimated for shorter tests. However, a reverse pattern was observed in the estimated standard errors of 97.5th percentiles in small and medium samples from a chi-squared ($df = 5$) distribution in all test targeting conditions. For small and medium samples, longer test seems to have larger standard error for the 97.5th percentiles. The reason for this occurring is unclear. Test targeting showed no effect on the standard error estimates. Sample distribution also showed no effect on the standard error estimates except 95th and

97.5th percentiles from a chi-squared ($df = 5$) distribution had a larger estimated standard error than that from the other three sample distributions.







Figure 4.3 The Estimated Standard Error of the Parameter Estimates from Simulation One

4.2 Simulation Two with Heterogeneous Populations

Simulation Two was carried out with data generated from heterogeneous populations in the data generating model. Three background variables were gender, LBOTE and Indigenous status and these variables were used as covariates in the data generation model (refer to Section 3.4 for data generation), and as conditioning variables for drawing plausible values. Values of these three background variables from a sample of 2,000 students, which were selected from the Northern Territory NAPLAN 2015 data, were used in the simulation. The simulations were carried out with one sample size of 2,000 students because the results from Simulation One showed that when the sample size was very large, e.g. 10,000, increasing the number of plausible values used in the estimation had little effect on the bias in the parameter estimates, the RMSE or the standard error of parameter estimates (see Section 4.1). With a small sample size of 200, the subgroups would have a sample size too small to be able to draw a meaningful conclusion. For instance, in Australia the proportion of Indigenous students is about 5.5% which would result in about 11 students from a sample size 200. A test length of 80 items was not included in this part of the simulations as in practice very long test is seldom used.

The generating parameters and the derived parameters for Simulations Two are listed in Table 3.3. The derived parameters, 90th and 97.5th percentiles and a cut score for ability levels, were estimated based on a normally distributed sample $N(0, 2.3896)$. These parameters were kept the same across the different test lengths and targeting conditions that were used across all simulation condition in Simulation Two. The generating regression coefficients are listed in Table 4.2. There were three regression coefficients associated with three independent variables which represent population background covariates: β_1 for gender with a small value, β_2 for LBOTE status with a medium value and β_3 with a relatively large value for Indigenous status.

Table 4.2 Generating Coefficient Values for Simulation Two

Generating Coefficient	Value	Background Variable
β_1	-0.09	Gender
β_2	-0.57	LBOTE Status
β_3	-1.00	Indigenous Status

The following sections present the results from Simulation Two. As was the case for Simulation One, the parameter estimates were aggregated into six sets and each set of results was based on a different number of plausible values used. Three regression coefficients and five population parameters (mean, variance, percentage in the top band, 90th and 97.5th percentiles) were estimated. The 95th percentile was not estimated for this part of the simulation as it had a very similar property to the 97.5th percentile in Simulation One. The bias and RMSE indices of estimated parameters, as well as estimated standard error were evaluated. The detailed results are given in Appendix 5. The graphic presentation of the bias and the RMSE of the parameter estimates are shown in Figure 4.4 to Figure 4.7. The estimated standard errors are presented in Figure 4.8 and Figure 4.9.

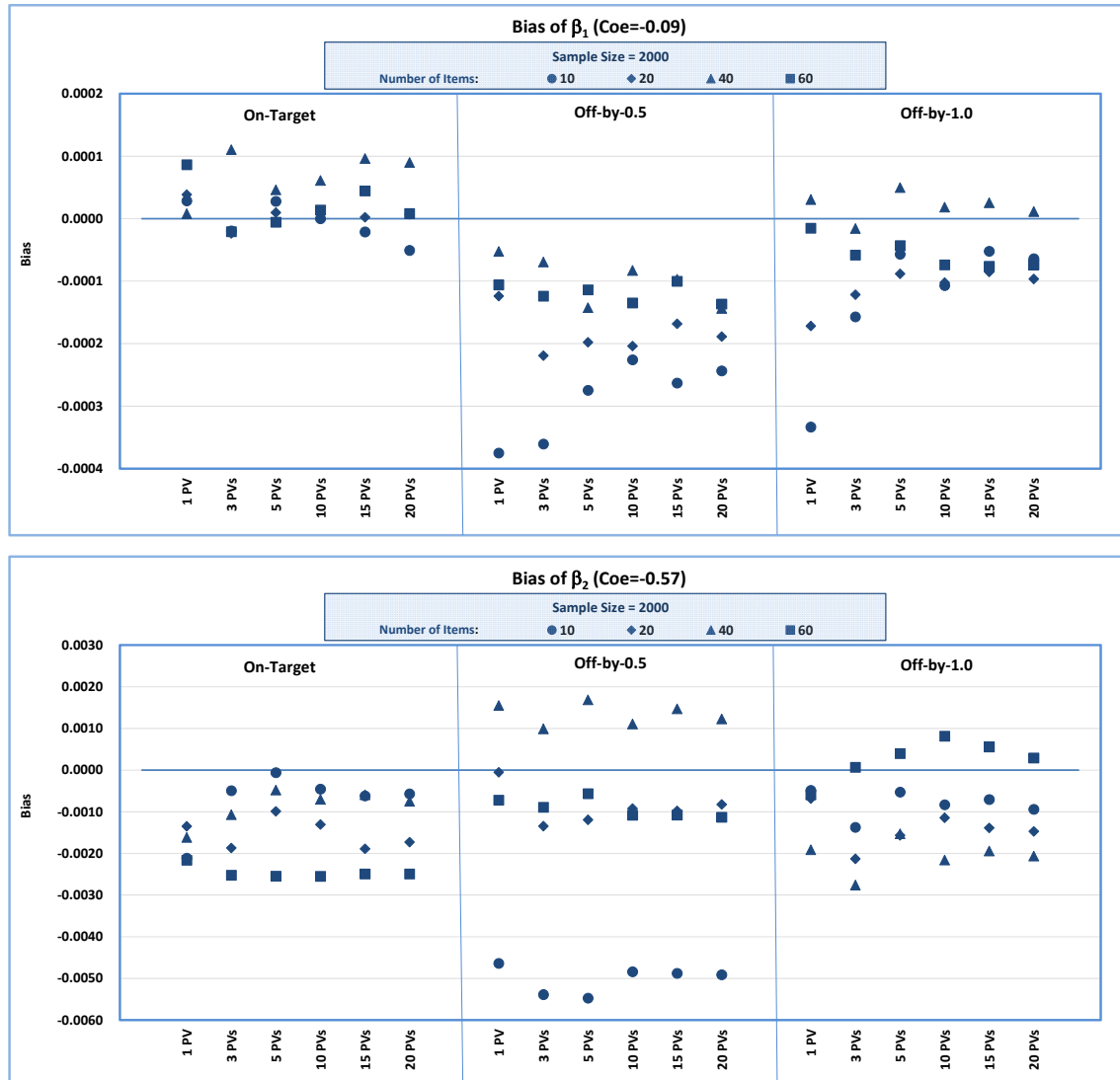
Figure 4.4 to Figure 4.7 show the bias in the estimated regression coefficients, the RMSEs of the estimated regression coefficients, the bias in the other parameter estimates and the RMSEs of the other parameter estimates, respectively. There are several plots in each of these figures, each plot depicts the outcome of one specific parameter estimate. There are three panels in each of these plots, each panel shows one test targeting condition: on-target, small off-target by 0.5 logits and medium off-target by 1.0 logit. Within each panel, four series of results have been plotted with a different symbol, one for each test length (10, 20, 40 and 60-item tests). The horizontal axis shows the number of plausible values used, and the vertical axis shows the bias or the RMSEs of a specific parameter estimates. The same symbol scheme that was used in the previous sections is used for Figure 4.4 to Figure 4.9.

4.2.1 The Bias and RMSEs of Regression Coefficient Estimates

Figure 4.4 shows the bias in three plots, one for each of the estimated regression coefficients ($\beta_j, j = 1, 2 \text{ or } 3$). Figure 4.4 shows that the bias in each of the estimated three regression coefficients using different numbers of plausible values was very small (between -0.0055 and 0.0017 logits), especially for the β_1 estimates (between -0.0004 and 0.0001 logits). It also shows that for all three regression coefficient estimates (β_j), there was no clear trend in estimated bias with increasing number of plausible values used, increasing test length, or being under different test targeting conditions.

Figure 4.5 shows the RMSEs in three plots, one for each of the regression coefficients estimates. It shows that the RMSEs for the three regression coefficient estimates were small and less than 0.0687 logits, even when using just one plausible value in the estimation. Similar to the RMSE of parameter estimates in Simulation One, Figure 4.5 shows that the RMSEs of the three regression coefficient estimates reduced with an increase in the number of plausible values used. On average, about 83% of the total reduction in the RMSE estimates (using 1 to 20 plausible values) occurred between using 1 and 5 plausible values. with further increasing the number of plausible values used from 5 to 20, the RMSEs only reduced by up to 0.0022 logits, or less than 18% of the total reduction. The RMSEs of estimated β_1 were the smallest and the RMSE of estimated β_2 were the highest among the three regression coefficients. When test lengths were increased, the RMSEs of the estimated regression coefficients reduced. However, the reductions in the RMSEs between the two short tests, from 10 to 20-item tests, were much larger than the reductions in the RMSEs between the two longer tests, from 20 to 40-item tests when other simulation conditions were kept the same. In other words, the magnitude of reduction in the RMSEs decreased with increases in the test lengths. For example, the reduction in the RMSEs of the estimated β_2 were 0.0083 logits or more when the test length increased from 10 to 20 items. But when the test length increased from 40 to 60 items, the reduction the RMSEs were less than 0.0028 logits while other simulation parameters were the same and the same number of plausible values were used. In fact, the RMSEs of the regression coefficient estimates from a 40-item test were very close to

the RMSEs estimated from a 60-item test. There was no effect observed when changing test targeting conditions on the RMSE of the regression coefficients estimates.



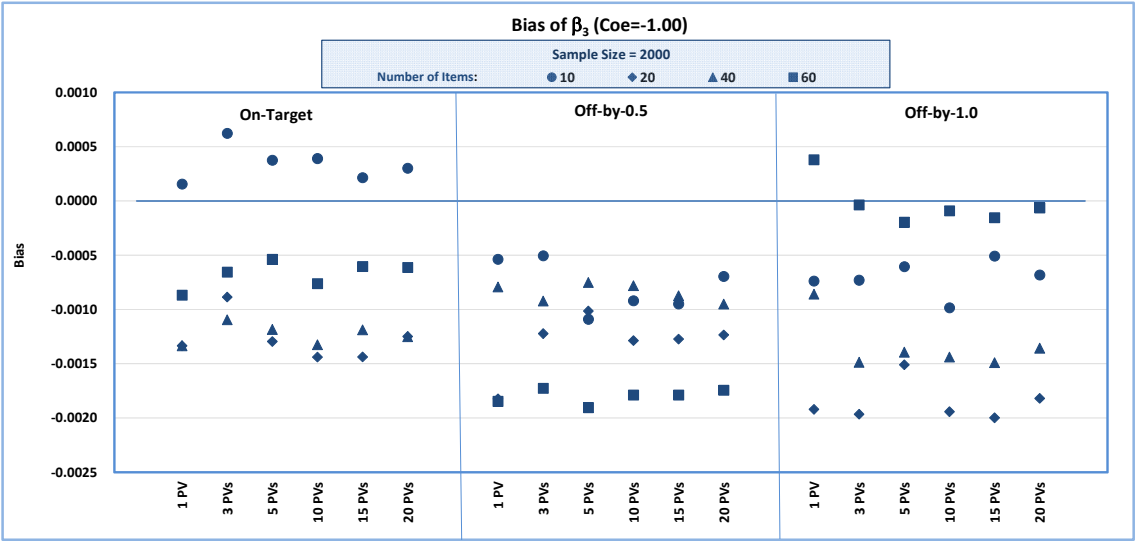
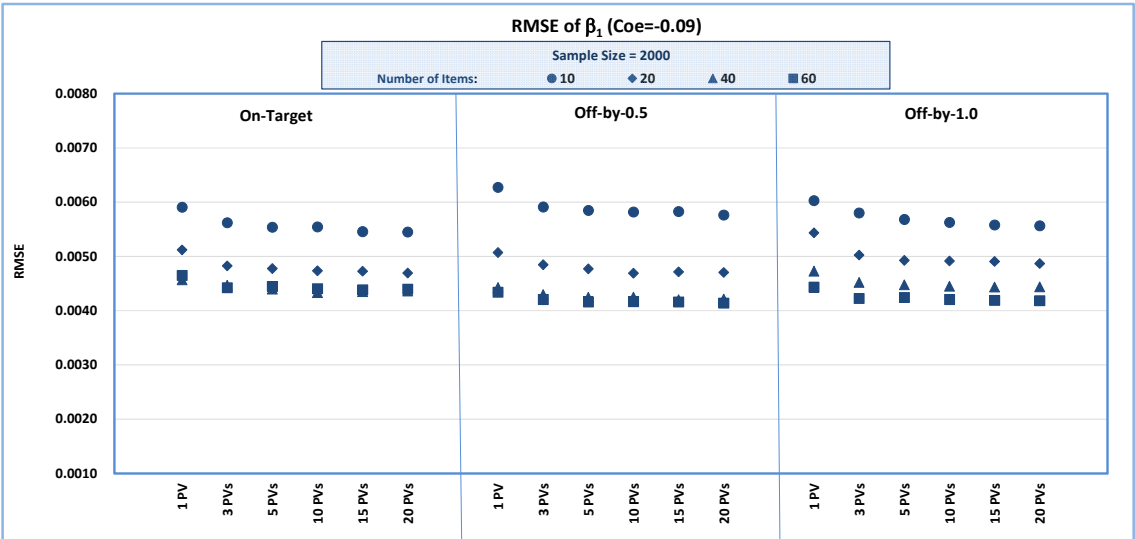


Figure 4.4 The Bias in the Estimated Regression Coefficients from Simulation Two



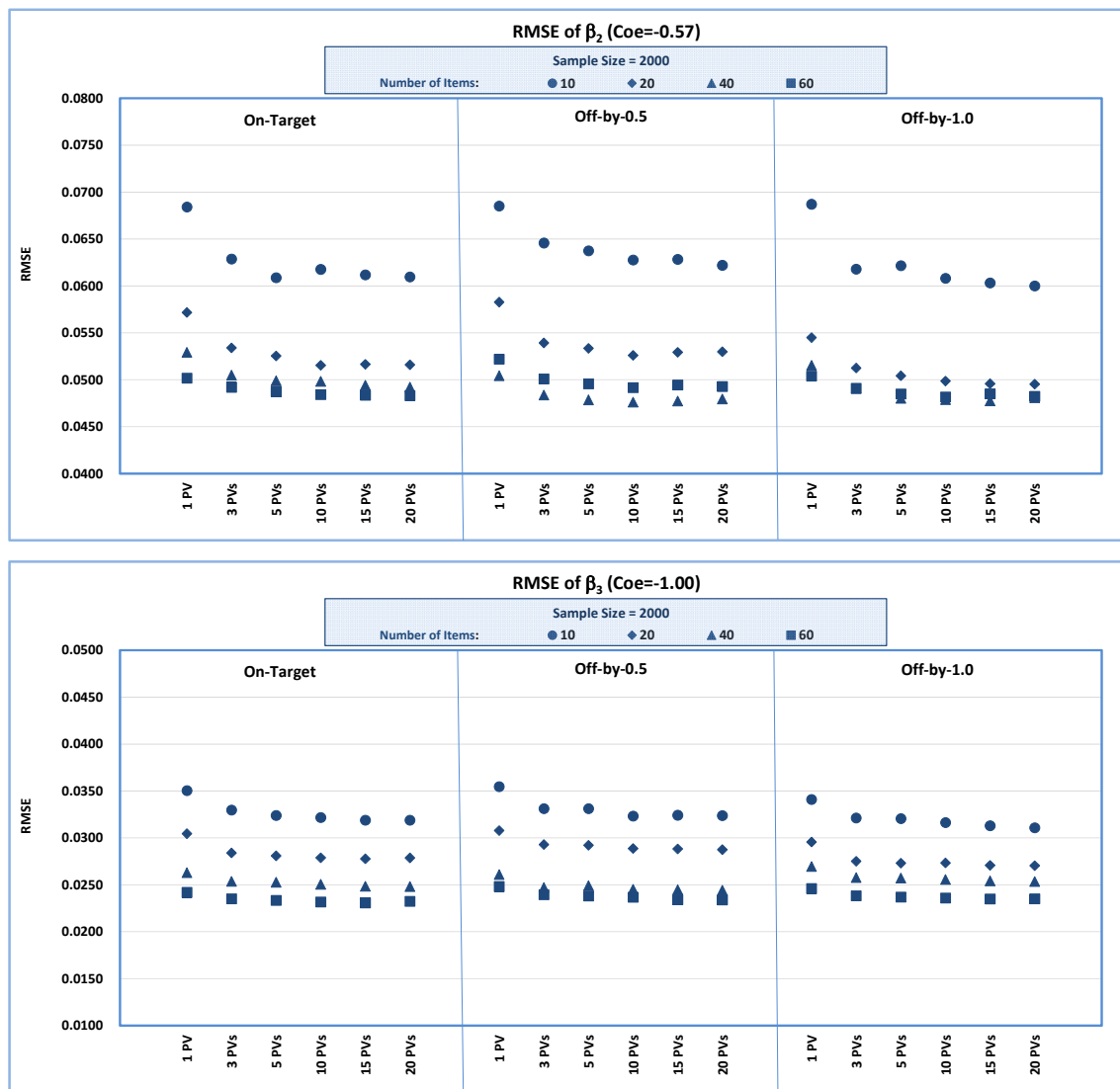
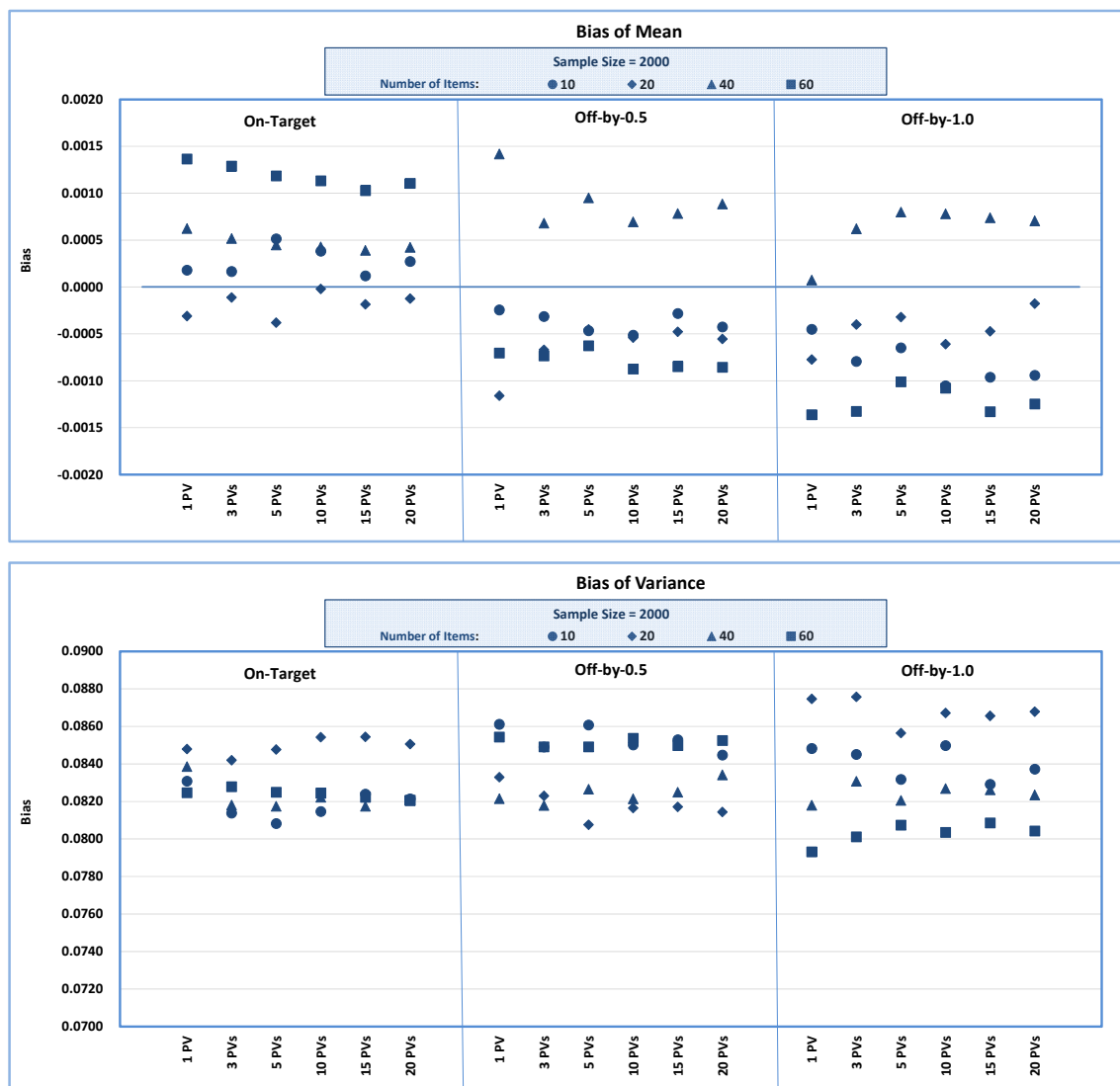


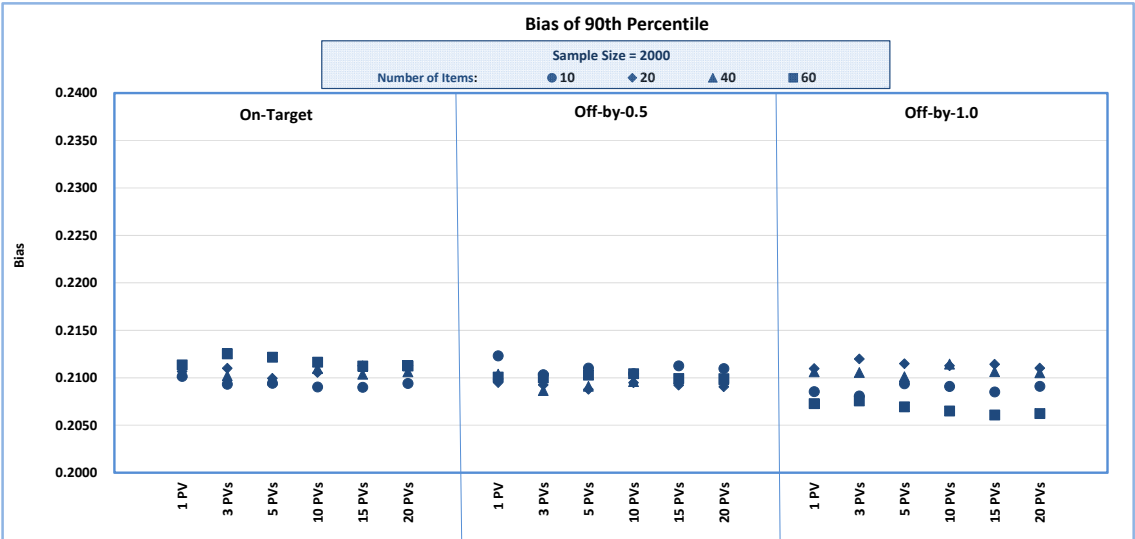
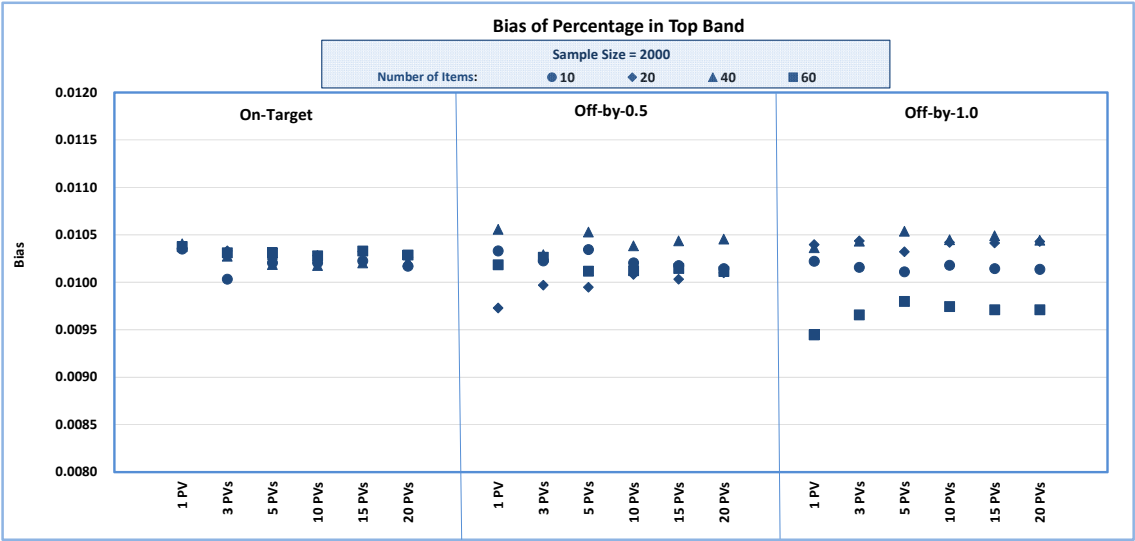
Figure 4.5 The RMSEs of the Estimated Regression Coefficients from Simulation Two

4.2.2 The Bias and RMSEs of the Other Parameter Estimates

Figure 4.6 presents the bias in five other parameter estimates (other than the regression coefficients) using different numbers of plausible values by simulation conditions. It was observed that there was a very small bias in the population mean estimates (± 0.0020 logits). With the increase in the number of plausible values used, there was no increasing or decreasing trend observed on the bias in the mean estimates. It was observed that there was a positive bias in the variance, the percentage in the top band and in the 90th and

97.5th percentile estimates across all simulated conditions (Figure 4.6). There was also no increasing or decreasing trend observed on the bias in the five parameters estimated with an increase in the number of items in a test or improved test targeting conditions. The magnitude of the bias in the estimated percentage of students in the top band was very similar across all 12 simulated conditions regardless of the number of plausible values used. They were all within the range of 0.0094 to 0.0110 logits. This indicated that neither the number of plausible values used nor the simulation conditions affected the bias in the percentage estimates in the top band. The bias in the two percentile estimates showed that the bias was larger with a higher percentile rank. The percentile estimates had the largest bias among the five population parameter estimates.





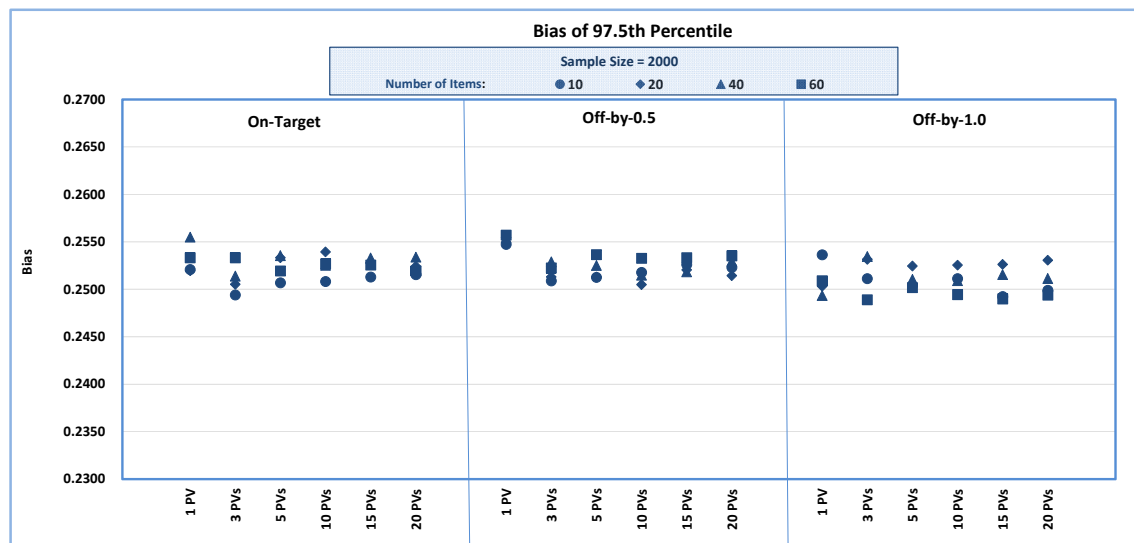
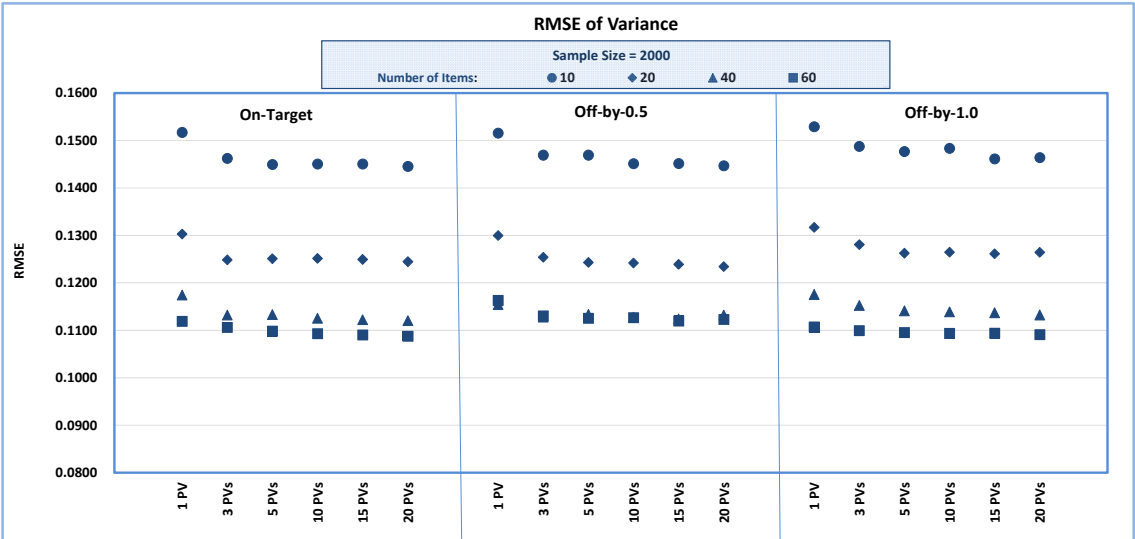
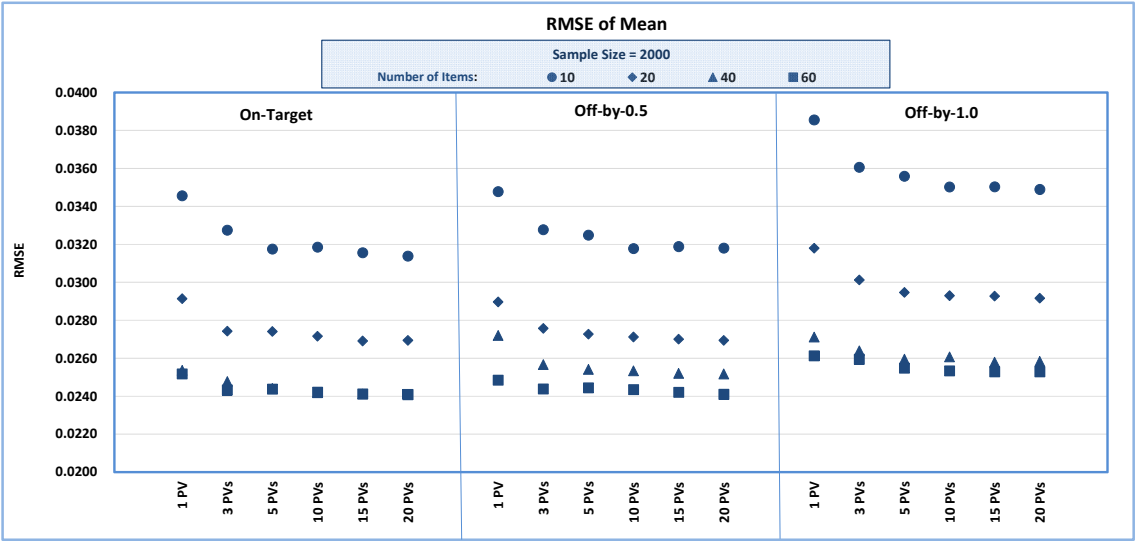


Figure 4.6 The Bias in the Other Parameter Estimates from Simulation Two

Figure 4.7 shows the RMSEs of the parameter estimates using different numbers of plausible values by the simulation conditions. There was a similar pattern observed across the RMSEs of the five parameter estimates. That is, the RMSEs of the parameter estimates decreased with an increase in the number of plausible values used within each simulated condition. However, the size of the reduction was generally small. For example, the reductions in the RMSEs of the mean estimates were less than 0.0030 logits when the number of plausible values was increased from one to five. When the number of plausible values reached five, a further increase in the number of plausible values had very little effect on the RMSEs of the parameter estimates, less than 0.0007 logits difference. With an increase in the test length, the RMSEs of all five parameter estimates were reduced. However, when a test was longer than 40 items, there was very little change in the RMSEs of mean estimates and the RMSEs of variance estimates. With the improved test targeting status, the RMSEs of the five parameter estimates had a little difference.



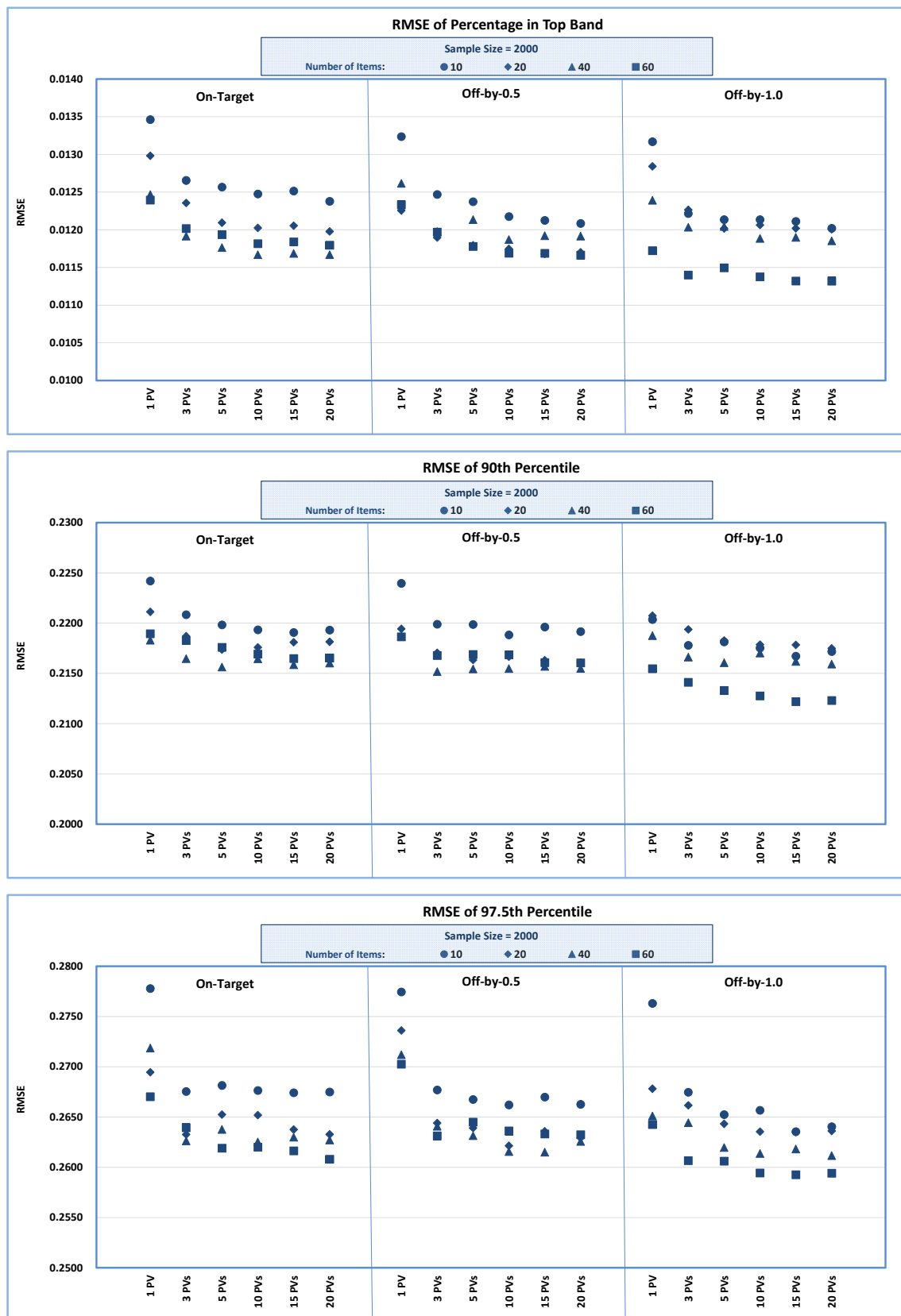
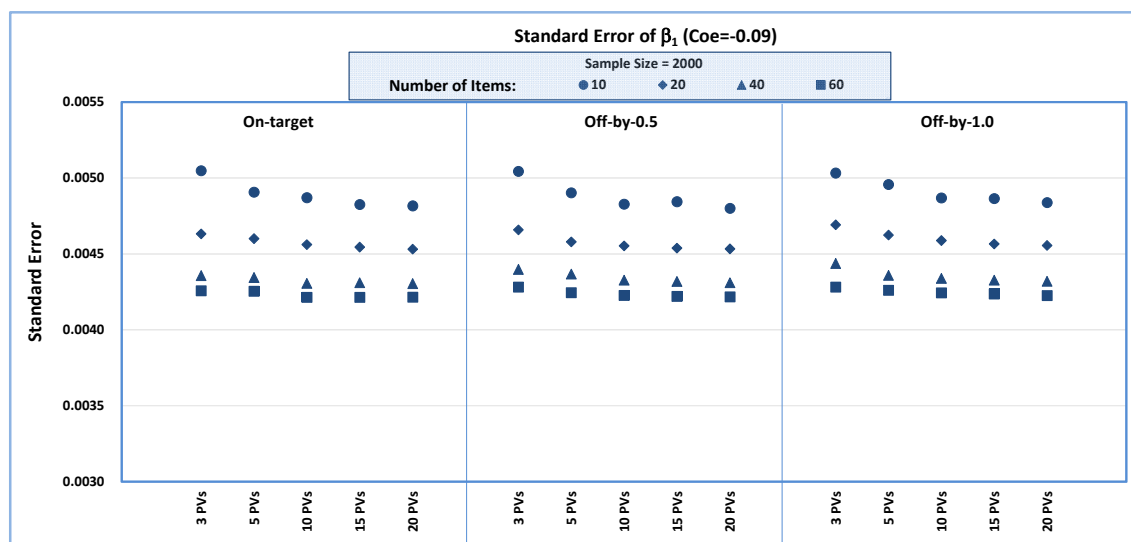


Figure 4.7 The RMSEs of the Other Parameter Estimates from Simulation Two

4.2.3 The Estimated Standard Error of Regression Coefficients

Figure 4.8 shows the estimated standard errors of three regression coefficients. A similar pattern was found in all three plots of standard errors. The estimated standard errors showed a decreasing trend with an increase in the number of plausible values used for all the simulated conditions. The estimated standard errors also reduced with an increase in the test length when the same number of plausible values was used, but the estimated standard errors of regression coefficients from 60-item tests were close to the estimated standard errors from 40-item tests. There was no clear pattern observed with changed test targeting conditions. The patterns observed in the estimated standard error and the RMSEs of the three regression coefficient estimates were very similar. The detailed standard error estimates are provided in Appendix 6.



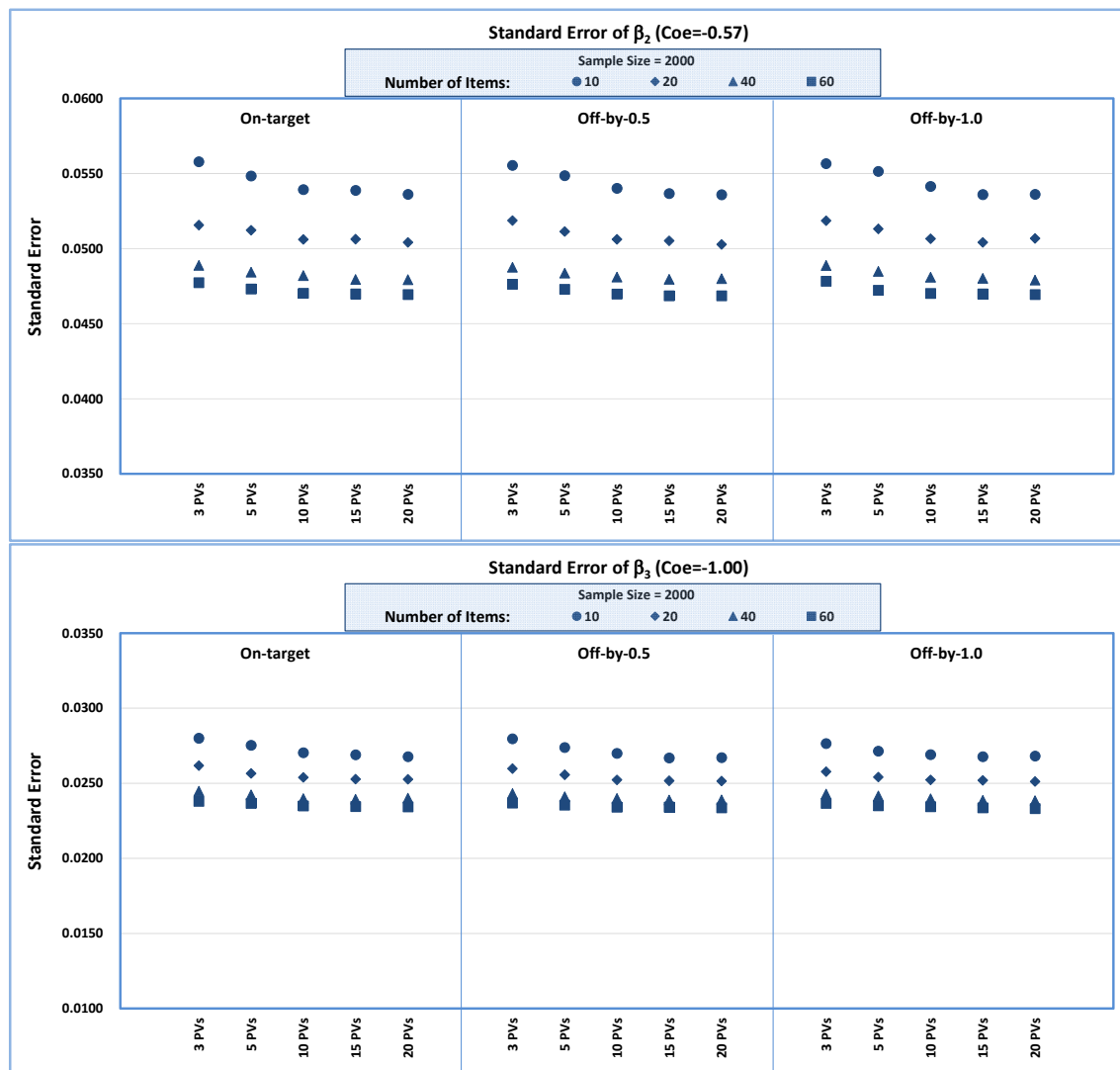
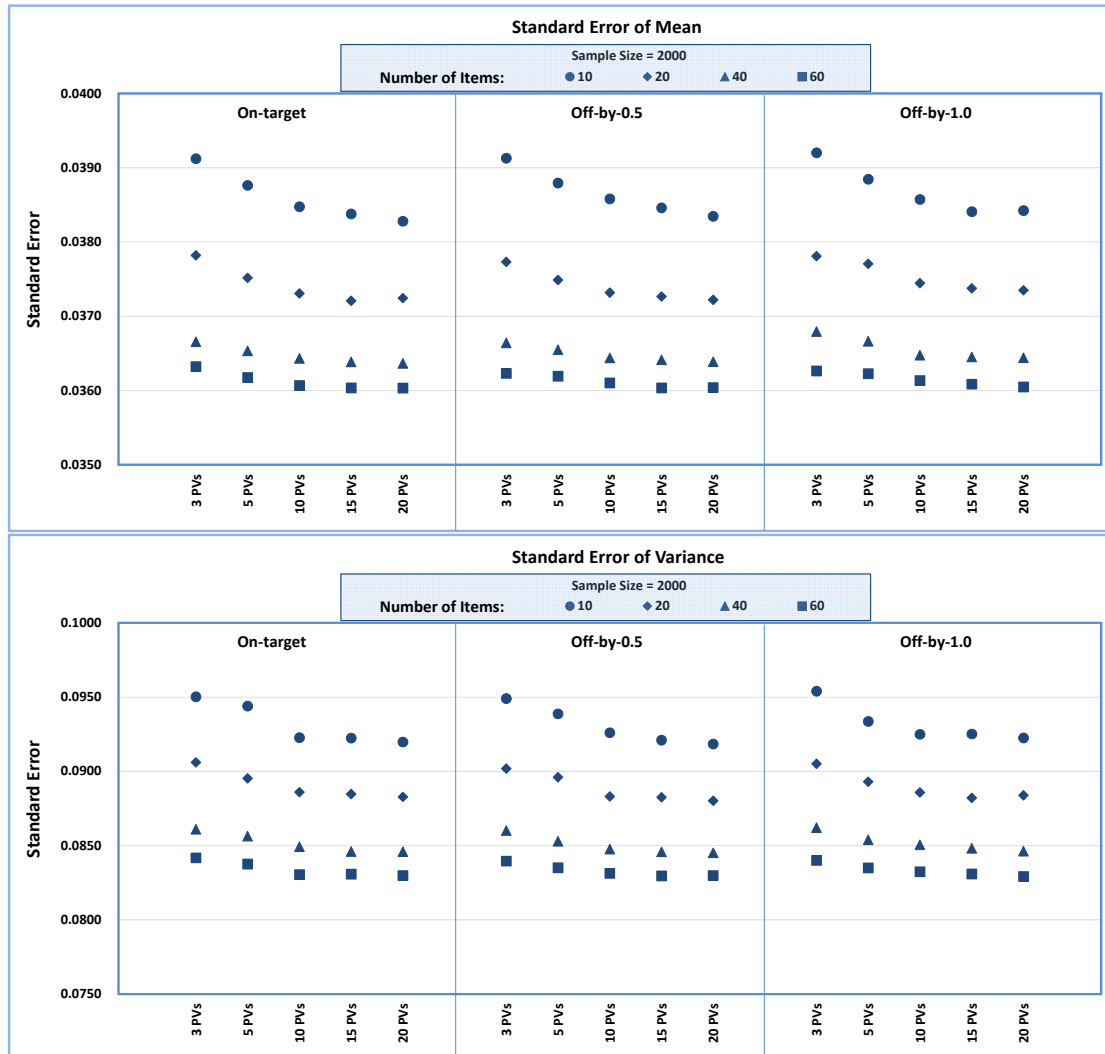


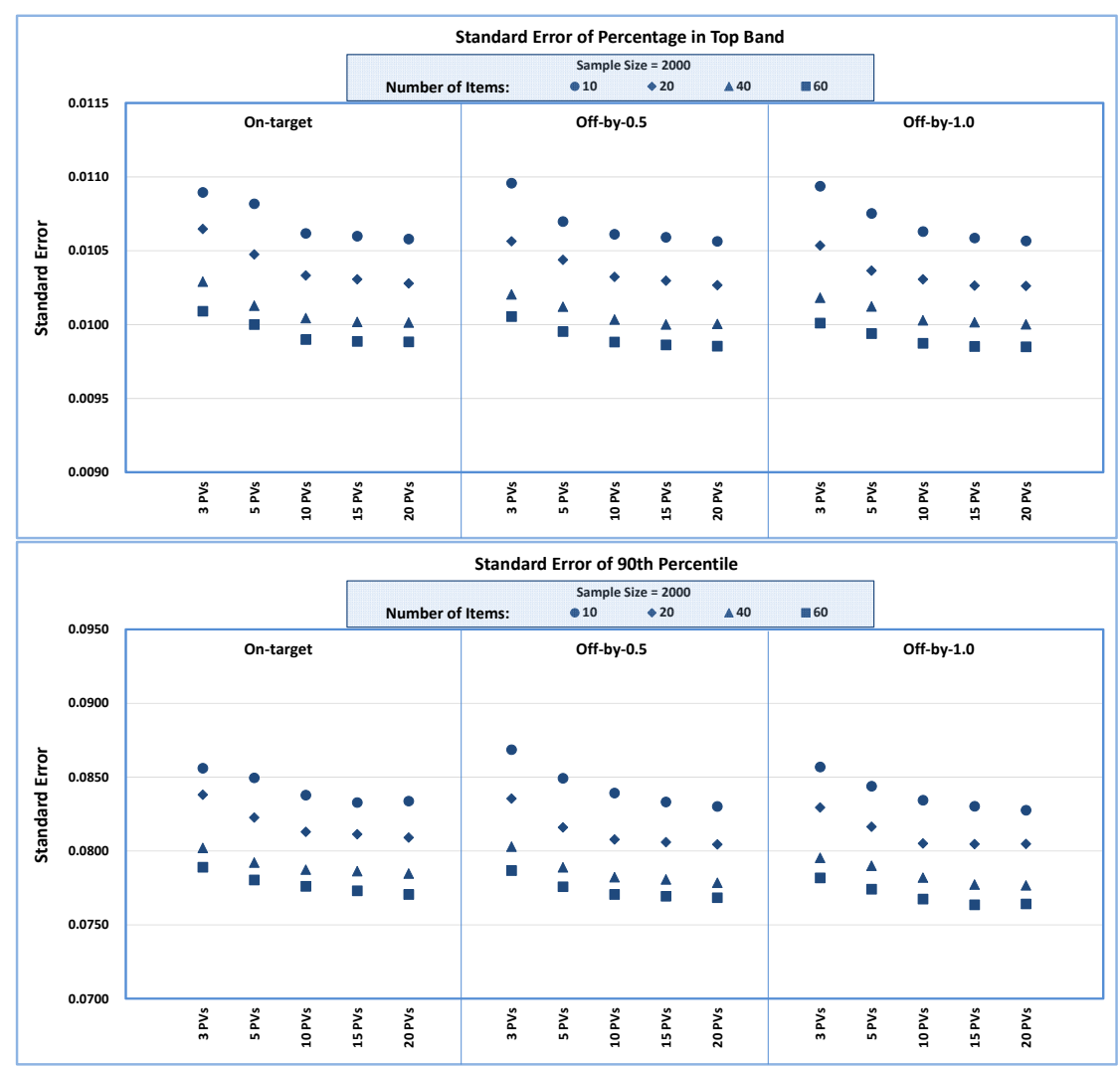
Figure 4.8 The Standard Error of the Regression Coefficients from Simulation Two

4.2.4 The Standard Errors of the Other Parameter Estimates

Figure 4.9 shows the standard errors of five parameter estimates (mean, variance, percentage in the top band, 90th percentile and 97.5 percentile) using different numbers of plausible values by simulation conditions. It shows that the estimated standard errors reduced with an increase in the number of plausible values used. The reduction in the estimated standard errors were more evident when the number of plausible values used were increased from three to five. The estimated standard errors also reduced when the

tests became longer. The magnitude of the reduction in the estimated standard errors became smaller with increases in test length. The estimated standard errors with different test targeting conditions were similar with the same test length using the same number of plausible values.





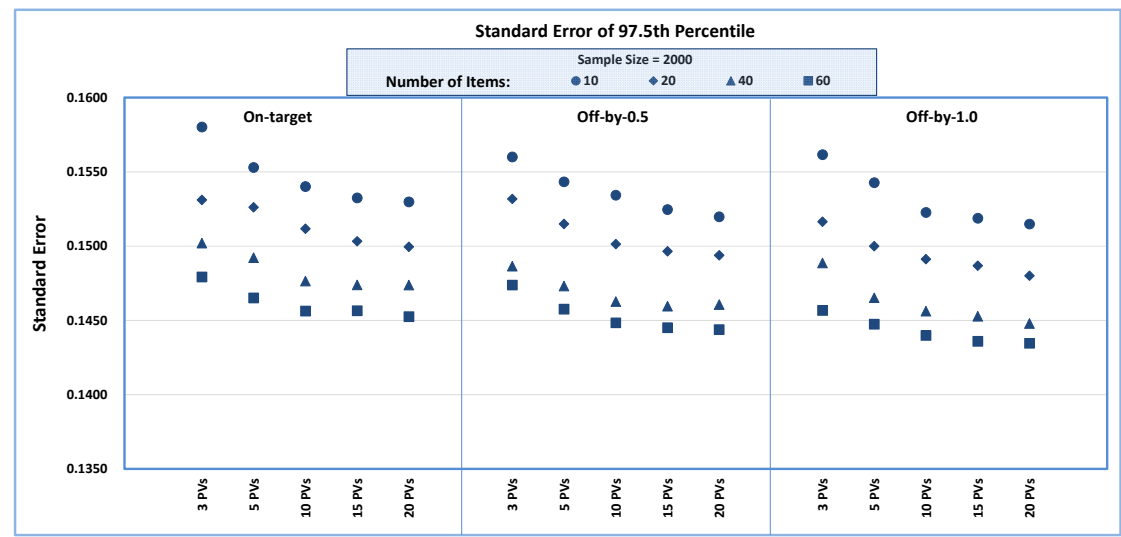


Figure 4.9 The Standard Error of Five Parameters Estimates from Simulation Two

Overall, Simulation Two has shown similar patterns to those that were observed in Simulation One when the same simulation condition was used. This is not a surprise result since Simulation One can be viewed as a special case of Simulation Two with $\beta_y = 0$ ($y = 1, 2$, or 3). Table 4.3 lists the findings from the two simulation studies. A summary of the findings is provided in Chapter 6.

Table 4.3 Summary of the Simulation Results Relating to the Research Questions

Research Questions		Simulation One			Simulation Two		
		Effect on the Bias	Effect on the RMSE	Effect on the Standard Error	Effect on the Bias	Effect on the RMSE	Effect on the Standard Error
Q1 – Increasing the number of PVs		No	Reduced	Reduced with small or medium samples	No trend	Reduced	Reduced
Q2 – Test Reliability	Increasing Test Length	Reduced, with exceptions	Reduced	Reduced with small or medium samples	No trend	Reduced	Reduced
	Test Targeting Status	No, except the bias in the mean estimates increased with the deterioration of the test targeting status	Mixed	No	No trend	No clear	Not clear
Q3 – Different Prior Distribution Assumptions		Not clear, except the bias was the smallest with normal distribution	No	No	N/A		
Q4 – Increasing Sample Size		Mean and variance reduced, not other four parameters	Reduced	Reduced	N/A		

Chapter 5 Results of Empirical Data Studies

In this chapter, the effect of the number of plausible values used on population parameter estimates was further explored using NAPLAN data. When it is possible, the findings from the empirical data studies were compared with the findings from the simulation studies. The empirical studies were carried out using the NAPLAN 2015 Numeracy data at four year levels: Year 3, 5, 7 and 9. The results were evaluated in two studies. Empirical Study One was carried out using all NAPLAN 2015 students, and Empirical Study Two was carried out with a subset of the data consisting of Indigenous students only, which has a much smaller sample size. The dataset consisted of data from eight jurisdictions: New South Wales (NSW), Victoria (Vic), Queensland (Qld), Western Australia (WA), South Australia (SA), Tasmania (Tas), the Australia Capital Territory (ACT) and the Northern Territory (NT). The outcomes were presented for all students in Australia (Aus) and by jurisdictions at each year level.

A total of 100 plausible values were drawn for each student. The population parameters were first estimated with each of 100 plausible values across all students and then the outcomes were combined into seven groups of results. Each group of results was estimated with a different number of plausible values, containing 1, 3, 5, 10, 15, 20 and 46 plausible values, respectively. The first six groups were combined in the same way as in the two simulation studies, and the seventh group consisted of 46, from the 55th to the 100th plausible values. Unlike the simulation study, there was no generating value (true value) to be used for estimating the bias or the RMSE indices in the empirical data study. In this part of study, therefore, the third group of results, which was estimated with five plausible values, were considered as the reference values. This is because the NAPLAN results are estimated using five plausible values like many other large-scale assessments. In addition, the findings of the simulation studies revealed when the number of plausible values increased to more than five, the effects of increasing the number of plausible values on population parameter estimates were very small. Another reason to use the estimates from five plausible values as the reference values is that the values of parameter estimates

are not the focus of this study, rather the relationship between the parameter estimates and the number of plausible values used in the estimation is the focus of this dissertation.

Although NAPLAN tests comprise of five domains: Reading, Writing, Spelling, Grammar and Punctuation and Numeracy, this study was carried out with the NAPLAN 2015 Numeracy data to take advantage of these different test lengths at each year level as outlined in Chapter 3.

Each year, the item parameters of the NAPLAN tests are calibrated with the Rasch model and then subsequently equated to the established NAPLAN scale. Because the main concern of this dissertation is estimating population parameters using plausible values, not the estimation of item parameters, NAPLAN item parameters were not re-estimated for this study. The plausible values for this study were drawn using the item parameters published in the NAPLAN 2015 Technical Report Appendix C (ACARA, 2015b). The analysis of the results were based on the logit metrics rather than the NAPLAN scale scores.

Again, since all items in the NAPLAN 2015 Numeracy tests were dichotomously scored, the simple logistic model implemented in ACER ConQuest 4 (Adams *et al.* 2015) was used for drawing the plausible values in this study¹¹. The model is as specified in Equation 2.4. The NAPLAN 2015 data file was split by jurisdictions into eight data files first at each year level, then 100 plausible values per student were drawn by jurisdiction by year level separately, following the procedure described in the NAPLAN 2015 Technical Report (ACARA, 2015b). The conditioning variables were: gender, Indigenous status, LBOTE status, school sector, school geolocation, parent education,

¹¹ With this part of study, the plausible values were drawn with a unidimensional IRM model. For NAPLAN, five plausible values were drawn with a multi-dimensional IRM model for Reading, Numeracy, Grammar and Punctuation, and Spelling; and with unidimensional IRM model for Writing (ACARA, 2015b). Conditioning variables are included in the model.

parent occupation and schools' means (ACARA, 2015b). The ConQuest command file for the empirical data studies is given in Appendix 2.

The estimated population parameters were the same as those published in the NAPLAN National Report (ACARA, 2015a), they were: mean, variance, 20th and 80th percentiles, and standard errors were also estimated for each of these estimated parameters. Each year, the NAPLAN National Report reports the percentage of students in proficiency bands. There were six proficiency bands separated by five cut scores in each year level. For this study, only one arbitrary cut score per year level was used, and the percentage of scores above the cut score were estimated for each of the four year levels. The score point of the 80th percentile of Australian students achievement at each year level was used as the cut score. It was estimated with the assumption that Australian students achievement was normally distributed. Each year, the NAPLAN National Report not only reports the results for Australia and by jurisdictions, it also reports subgroup results. These subgroups were: gender, LBOTE and non-LBOTE students, Indigenous and non-Indigenous students, geolocation status, geolocation status by student Indigenous status, parents' education status and parents' occupation status (ACARA, 2015a). In this chapter, the effect of the number of plausible values used on the NAPLAN population parameter estimates are demonstrated with the two studies.

5.1 Results of Empirical Study One

This section presents the results from Empirical Study One (whole cohort in Australia and by jurisdictions). First, each of the population parameters estimated from M ($M = 1, 3, 10, 15, 20$ and 46) of plausible values was compared with its reference value. The differences between the parameters estimated with M sets of plausible values and the reference value were plotted. The results are discussed for the eight jurisdictions as well as for the whole of Australia. The detailed outcomes from the NAPLAN 2015 Numeracy tests by year level for each of eight jurisdictions and for Australia overall by the number of plausible values used are given in Appendix 7.

Table 5.1 lists the mean in logit by year levels and by jurisdictions and for Australia overall students and for Indigenous students. Because the item parameters were centred at zero (that is the average of the item parameters equals zero), hence the mean ability of a jurisdiction students can be seen as an indication of the test targeting status for the jurisdiction. A positive value in Table 5.1 indicates that the students had a higher average ability compared to the average item difficulties, therefore the test was on the easy side for the students. In comparison, a negative value indicates students had a lower average ability than the average item difficulties, therefore the test was hard for the students. More importantly, a mean that was close to zero indicates that the test was targeted well for the students, and a large mean, either positive or negative, indicates the test was off-target for the students. That is, the larger the absolute value of the mean, the worse the targeting of the test was.

Table 5.1 reveals that the test targeting status was quite different among the eight jurisdictions. For example, on average Year 3 Numeracy was well targeted for South Australian students, slightly easy for Victorian and the Australia Capital Territory students, and appeared quite difficult for the Northern Territory students. The difference in test targeting for the Year 3 Numeracy test among eight jurisdictions was about 1 logit. A similar difference in the test targeting range was found for the other three year level tests. For Indigenous students, the Numeracy tests were on the difficult side. The tests were the most difficult for the Northern Territory Indigenous students, with the test being up to 2 logits off-target for the Northern Territory Indigenous students.

Table 5.1 Test target Conditions by Year Level and by Jurisdictions and Australia

Jurisdictions and Australia	Year Level - All Students				Year Level - Indigenous Students			
	3	5	7	9	3	5	7	9
NSW	0.235	0.387	-0.035	-0.248	-0.520	-0.578	-0.978	-1.241
Vic	0.346	0.486	-0.013	-0.236	-0.265	-0.386	-0.834	-1.038
Qld	0.116	0.195	-0.165	-0.439	-0.582	-0.687	-0.955	-1.206
WA	0.073	0.174	-0.173	-0.251	-0.980	-1.090	-1.269	-1.385
SA	-0.012	0.054	-0.264	-0.539	-0.903	-0.998	-1.109	-1.389
Tas	0.104	0.141	-0.329	-0.566	-0.324	-0.396	-0.837	-1.074
ACT	0.348	0.466	0.005	-0.189	-0.421	-0.495	-0.796	-1.036
NT	-0.662	-0.768	-1.040	-1.194	-1.469	-1.772	-1.973	-2.016
Aus	0.190	0.307	-0.105	-0.322	-0.684	-0.797	-1.096	-1.307

The Parameter Estimates

Figure 5.1 to Figure 5.5 show the differences of the NAPLAN 2015 cohort parameter estimates between that estimated with M plausible values and the corresponding reference values. Each of the five figures shows differences for one of the five parameter estimates, respectively. Each figure has four plots, one for each year level. The results are arranged by Australia (Aus) on the left, followed by jurisdictions results arranged by the number of students (from the largest to the smallest) in each plot. The differences of the parameter estimates between those estimated with M plausible values and the reference values are plotted as a dot-line graphs, and the number of students are shown as a bar graph on the same plot. The number of students ranges from about 3,000 in the Northern Territory to about 90,000 in New South Wales, and about 300,000 for the whole of Australia in each of the four year levels.

Figure 5.1 to Figure 5.5 show that there was no increasing or decreasing trend observed with an increase in the number of plausible values used on any of the five parameter

estimates. The differences in these parameter estimates were within ± 0.025 logits for the mean, standard deviation, 20th and 80th percentiles, and were less than 0.5% for the percentage of students in the top band. When the number of students was about 60,000 or more, as they were in Australia overall and in New South Wales, Victoria and Queensland, the estimates from one plausible value to the estimates from 46 plausible values had very small differences. The estimates using the different number of plausible values were less than 0.010 logits for the mean, standard deviation, and 20th and 80th percentiles, and were less than 0.15% for the percentage of students in the top band. When the number of students was small, like Tasmania, the Australia Capital Territory and the Northern Territory, the differences of the parameter estimates using different numbers of plausible values increased. It also showed that the differences in the parameter estimates between using three or fewer plausible values and the reference value were larger compared to the differences between using 10 or more plausible values and the reference value. The patterns were similar across the four year levels, this indicated the test length had a limited effect if any on these results.

Examining the results from Victoria, the Australia Capital Territory and the Northern Territory, it could be argued that the test targeting had no effect on the parameter estimates. Victoria had a very similar test targeting status to the Australia Capital Territory's, but a much larger sample size, and the differences in parameter estimates across different numbers of plausible values used for Victoria were very small, but relatively large for the Australia Capital Territory. On the other hand, the Australia Capital Territory and the Northern Territory had very different test targeting statuses, but had similar sample sizes. The variations in parameter estimates using different numbers of plausible values for the Australia Capital Territory and the Northern Territory were both relatively large. This agrees with the findings from the simulation studies, that is, the test targeting has no effect on the parameter estimates regardless of the number of plausible values used in the estimation.



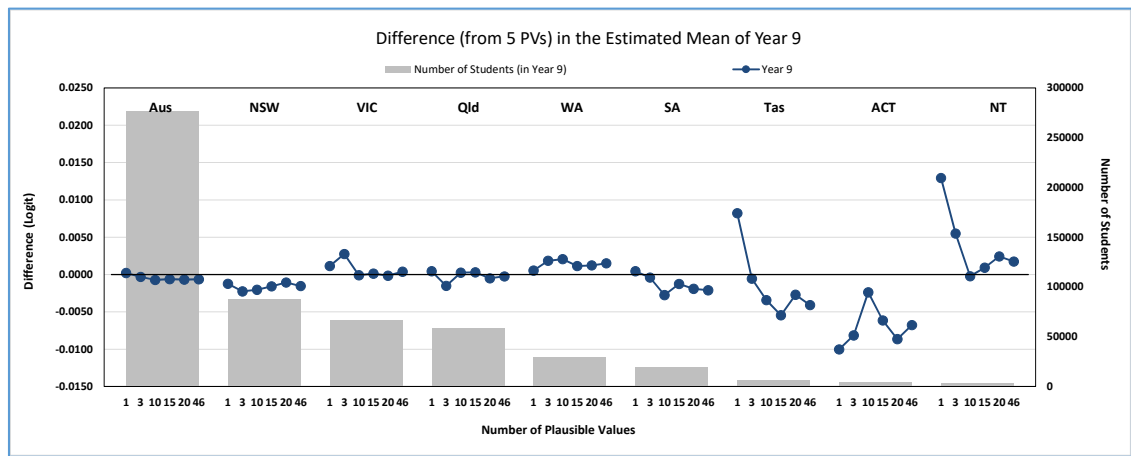
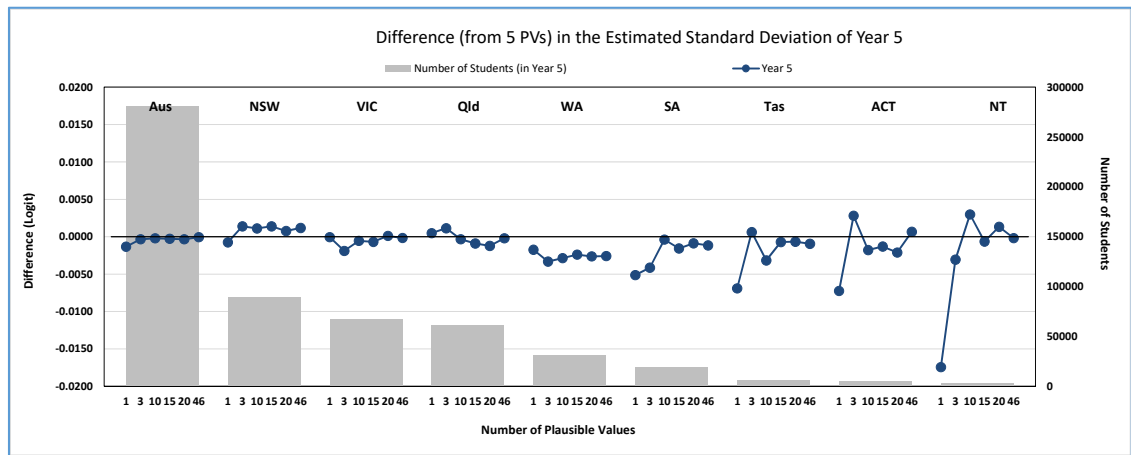
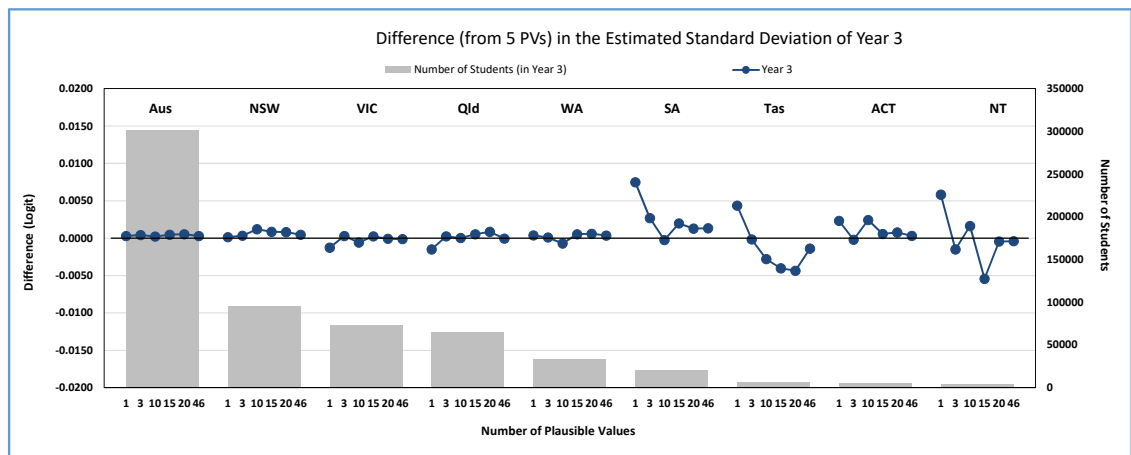


Figure 5.1 The Differences in the Mean Estimates in Empirical Study One



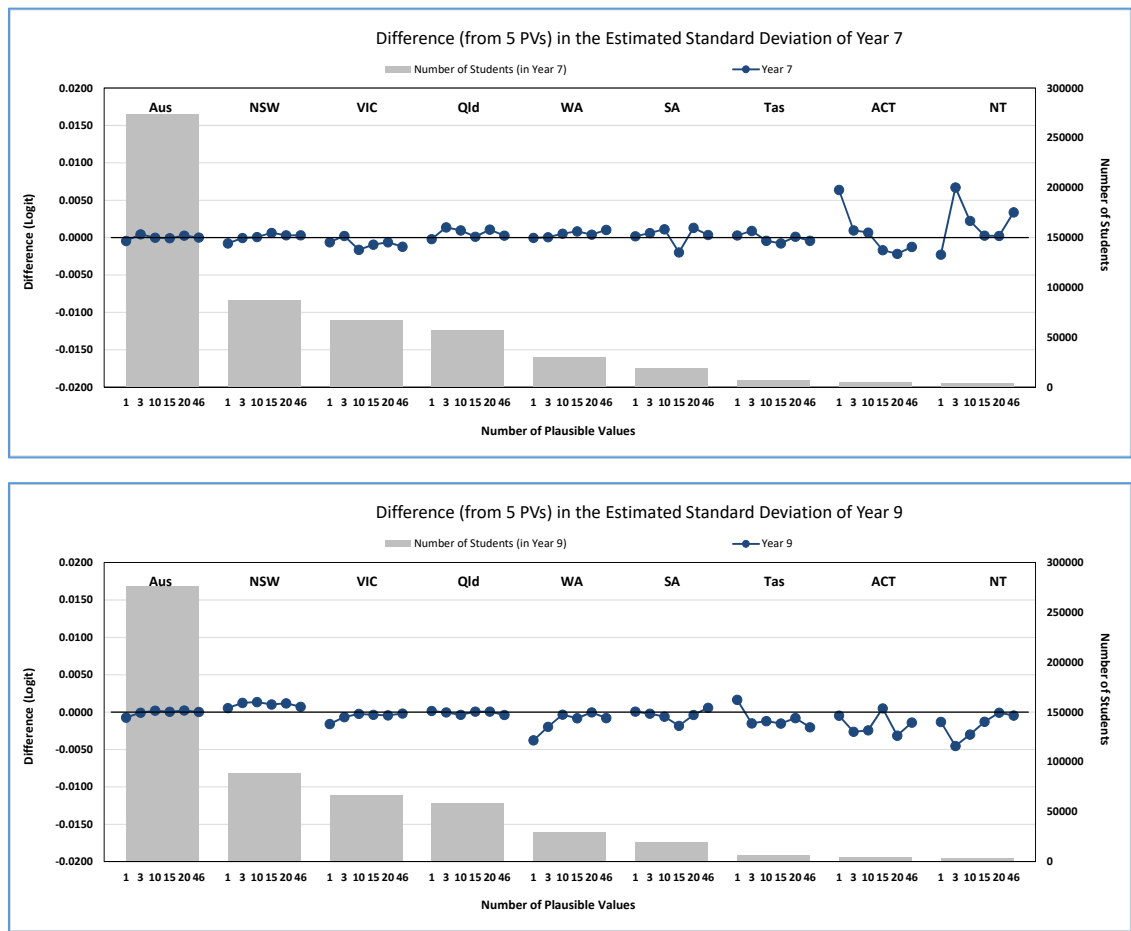
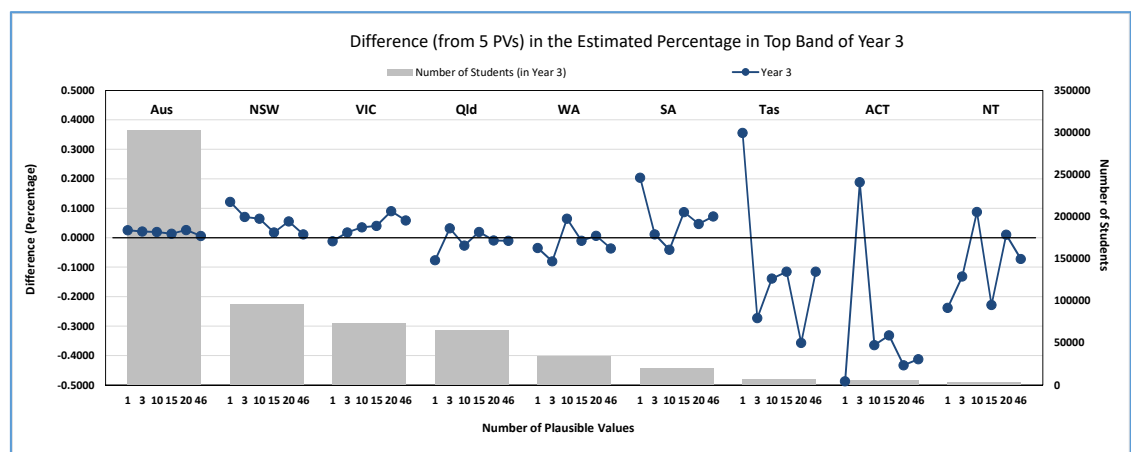


Figure 5.2 The Differences in the Standard Deviation Estimates in Empirical Study One



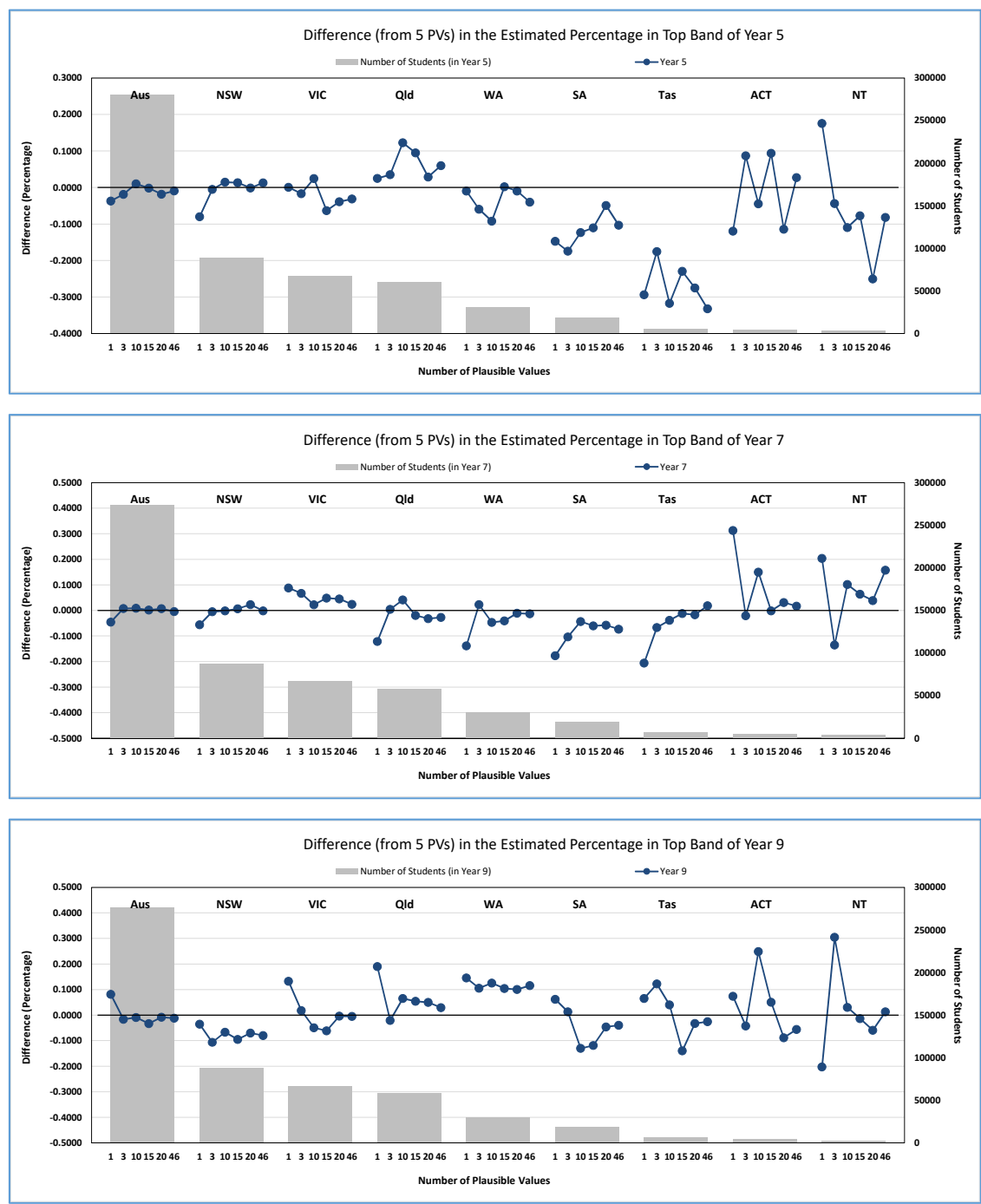
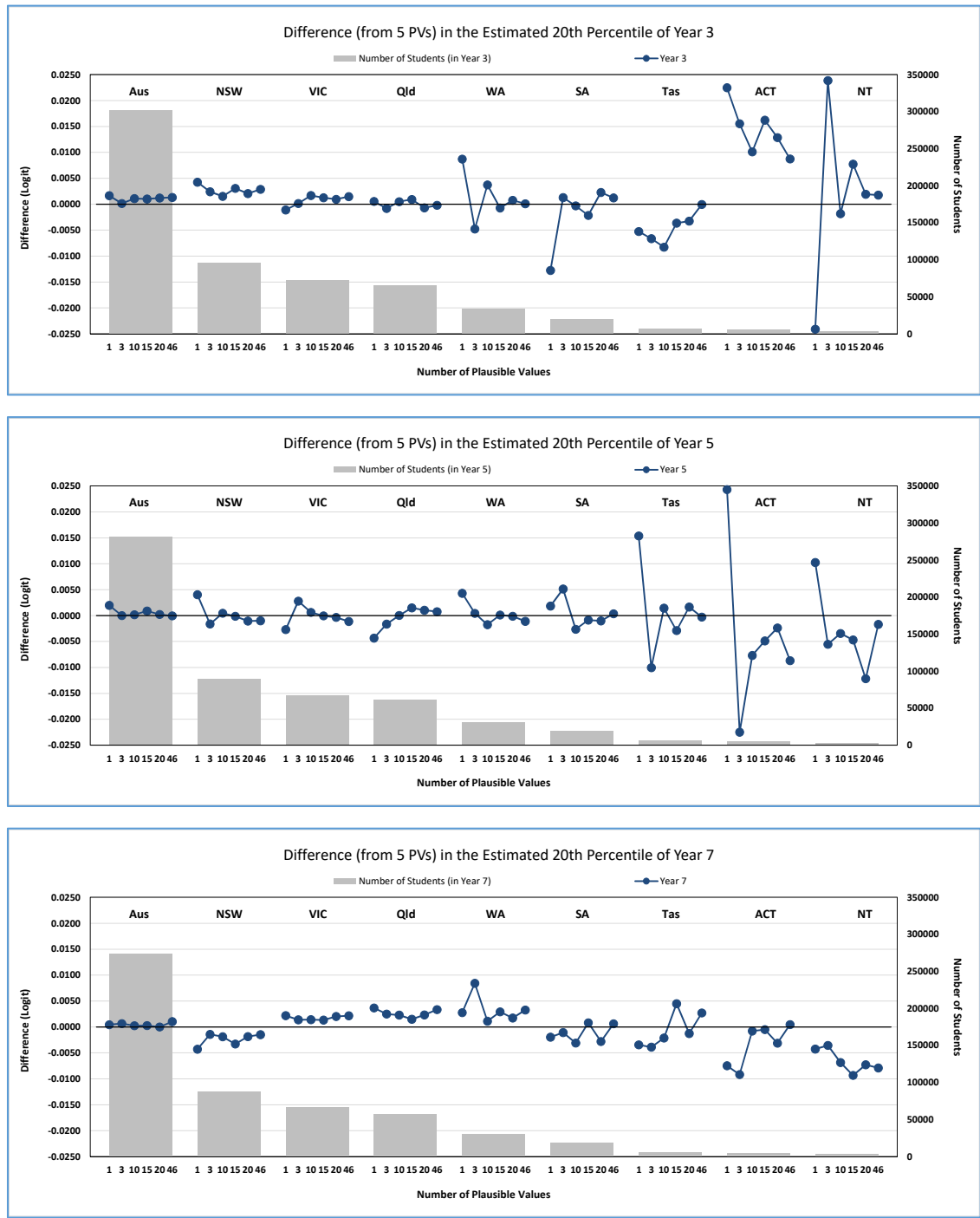


Figure 5.3 The Differences in the Percentage in the Top Band Estimates in Empirical Study One



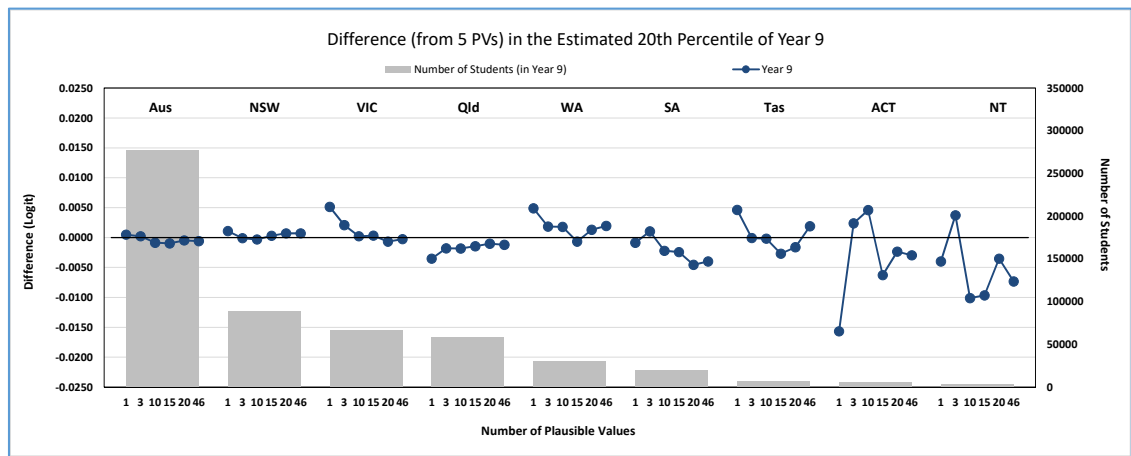
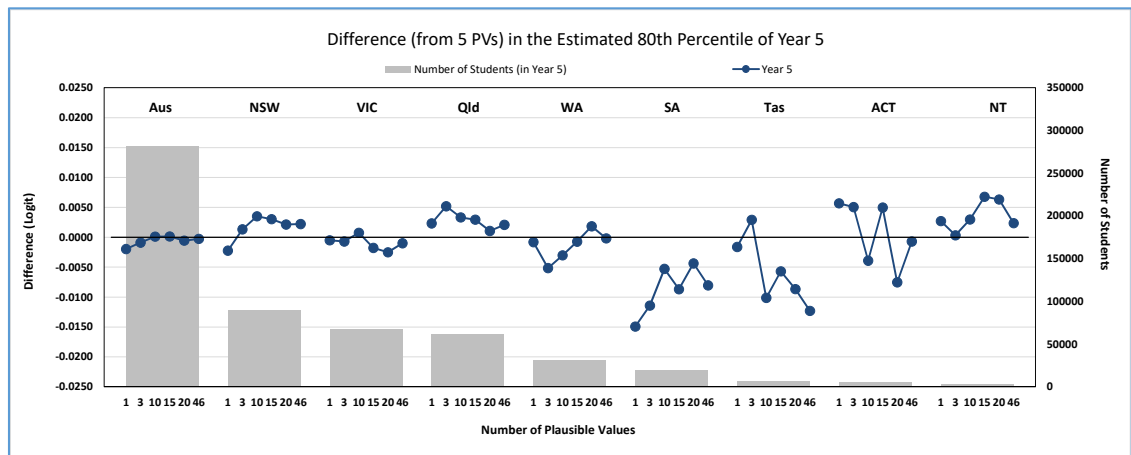
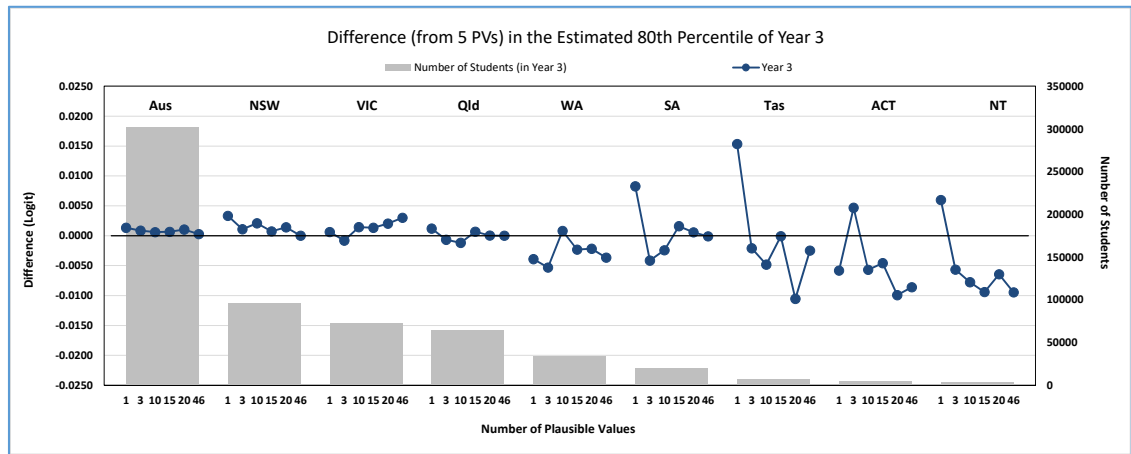


Figure 5.4 The Differences in the 20th Percentile Estimates in Empirical Study One



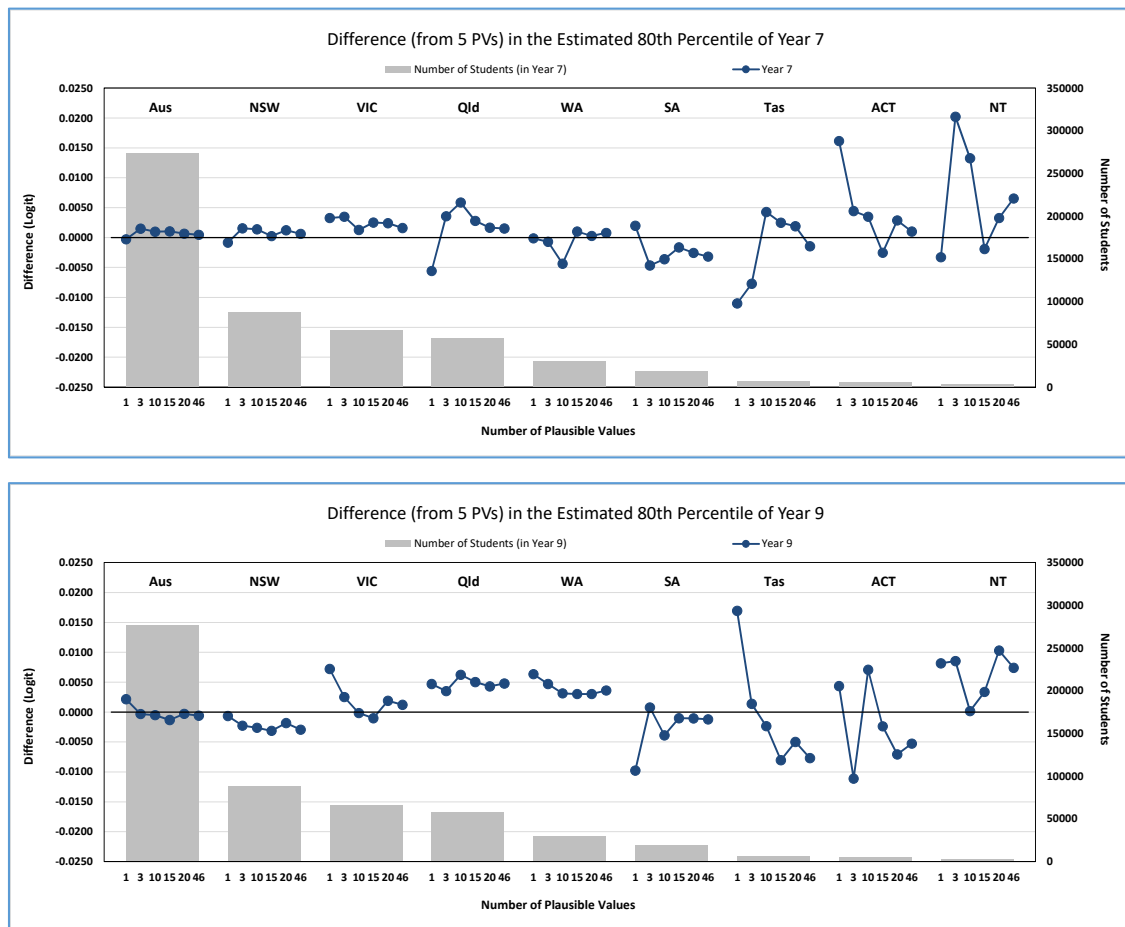


Figure 5.5 The Differences in the 80th Percentile Estimates in Empirical Study One

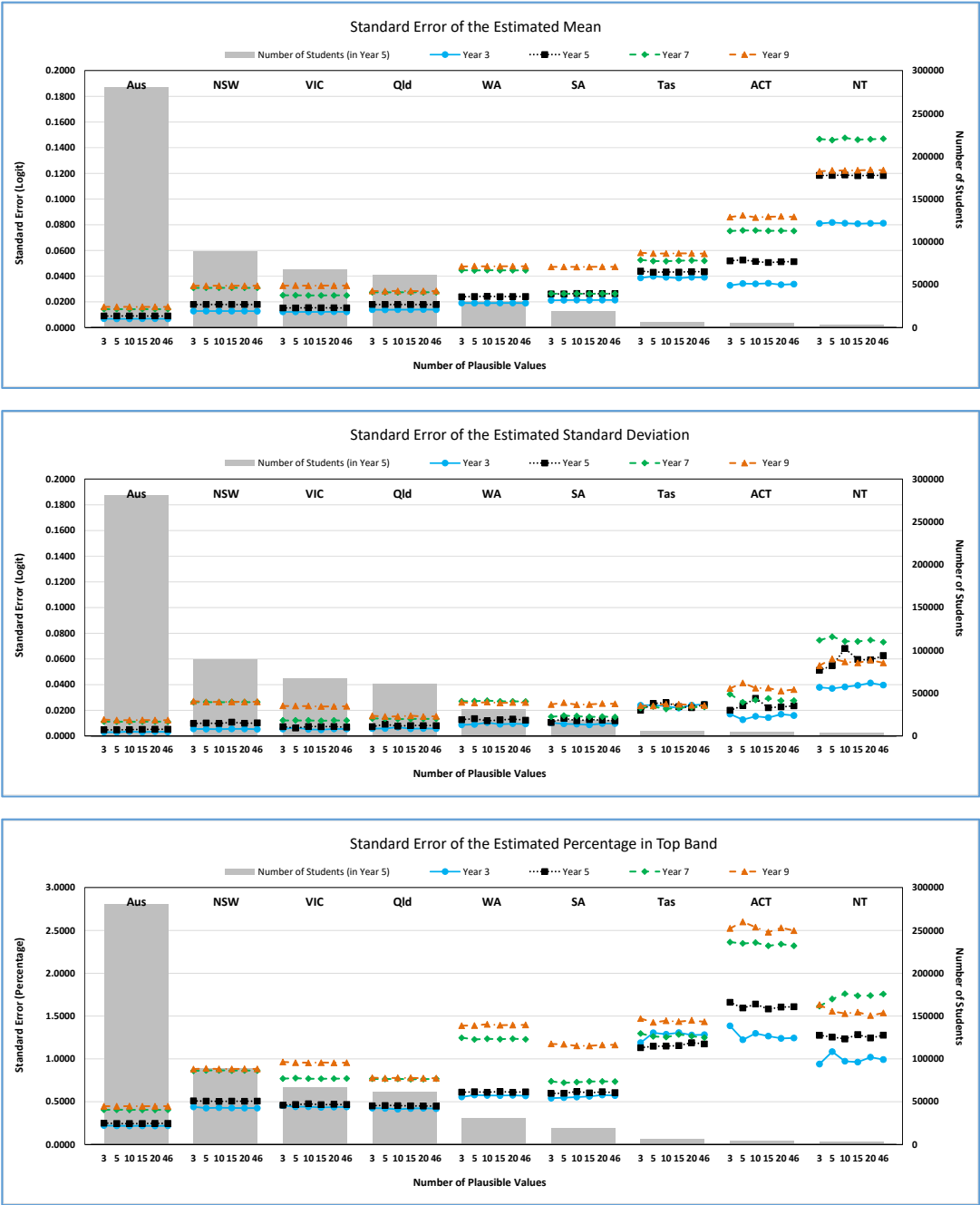
The Estimated Standard Errors

Figure 5.6 presents the standard errors of each parameter estimate using the different numbers of plausible values, from 3 to 46. Again, no standard error was estimated when using just one set of plausible value. There is one plot for each standard error of the parameter estimates. The standard errors in each plot are arranged by year level, by jurisdictions and for Australia overall. Australia's results are on the left, followed by the jurisdictions' results arranged from left to right in the order of sample size. Within each jurisdiction, the results are arranged by the number of plausible values used. The

estimated standard errors of parameter estimates are shown as dot-line graphs, the standard errors for each of the four year levels are presented by different symbols, and the sample sizes are shown as bars in each graph. As the number of students is similar across the four year levels for each of the jurisdictions and Australia, the number of Year 5 students are plotted to provide an indication of the approximate sample sizes.

Figure 5.6 shows that the standard errors of parameter estimates displayed little difference regardless of the number of plausible values used in the estimation by year level by jurisdictions. The differences in the standard errors for the mean, standard deviation, 20th and 80th percentile estimates using different numbers of plausible values were less than 0.0185 logits; and the differences in the standard error of the estimated percentages in the top band using different numbers of plausible values were less than 0.165%. The variation of estimated standard errors increased with a reduction in the number of students. There was no increasing nor decreasing trend observed in estimated standard errors with an increase in the number of plausible values used.

From the simulation studies, it was found that the standard error of parameter estimates reduced with increasing test length. However, this was not found with Empirical Study One. Among four year level tests, Year 7 and Year 9 had the same number of items (64) and were the longest tests, Year 3 had the shortest test. The standard errors of parameter estimates from Year 9 were often the largest and larger than the standard errors estimated from Year 7, whereas the standard errors of parameter estimates of Year 3 test were often the smallest among the four tests. This may be due to the number of students in Year 3 being the greatest among the four year levels and Year 9 being the least. No conclusion can be drawn on the effect of test length on the estimated standard error due to the confounding effect of test length and sample size.



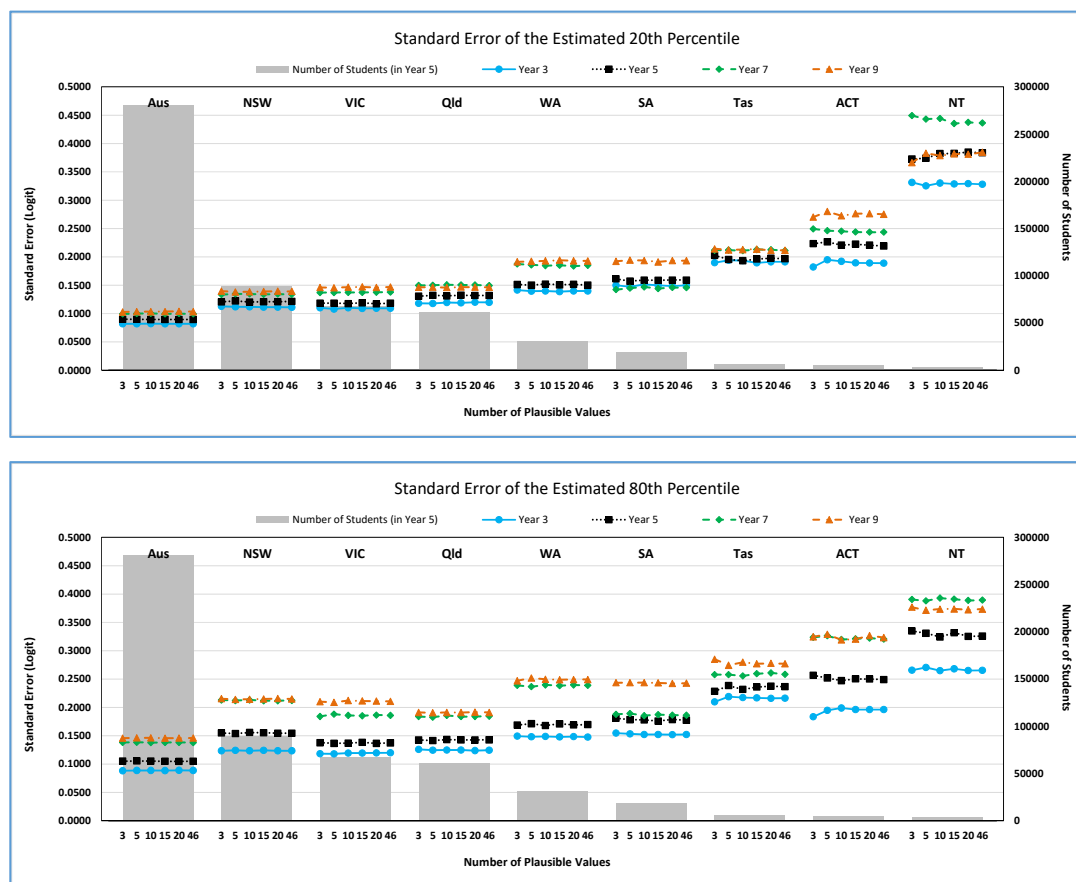


Figure 5.6 The Standard Errors of the Parameter Estimates in Empirical Study One

5.2 Results of Empirical Study Two

The previous section analysed the whole NAPLAN 2015 cohort data where each of eight jurisdictions had a relatively large number of students (2,957 or more). From the simulation studies, it was found that the sample size had the largest effect on population parameter estimates. In order to investigate the effect of the number of plausible values used on the population parameter estimates when sample sizes were small, this section presents the results that were evaluated with Indigenous students only. The number of Indigenous students in each jurisdiction varies between 121 (ACT) and 5,244 (NSW).

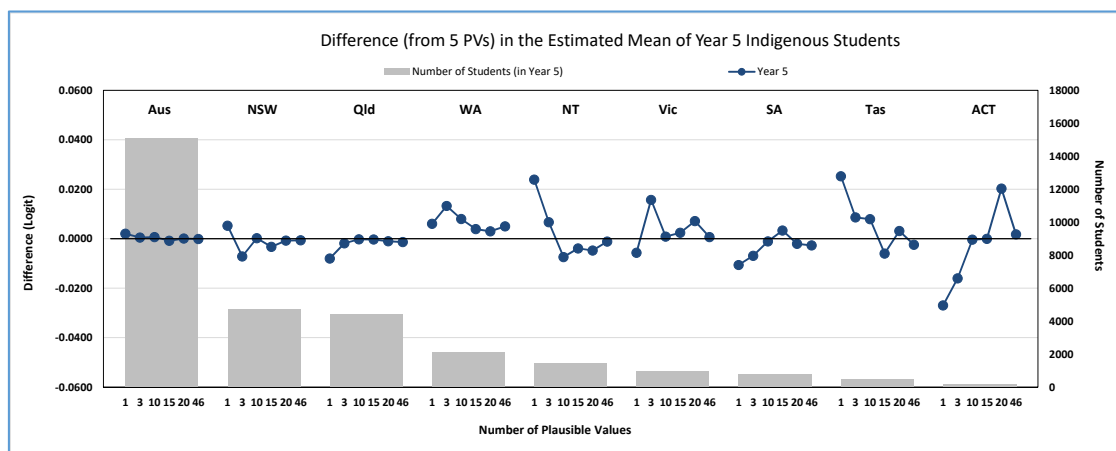
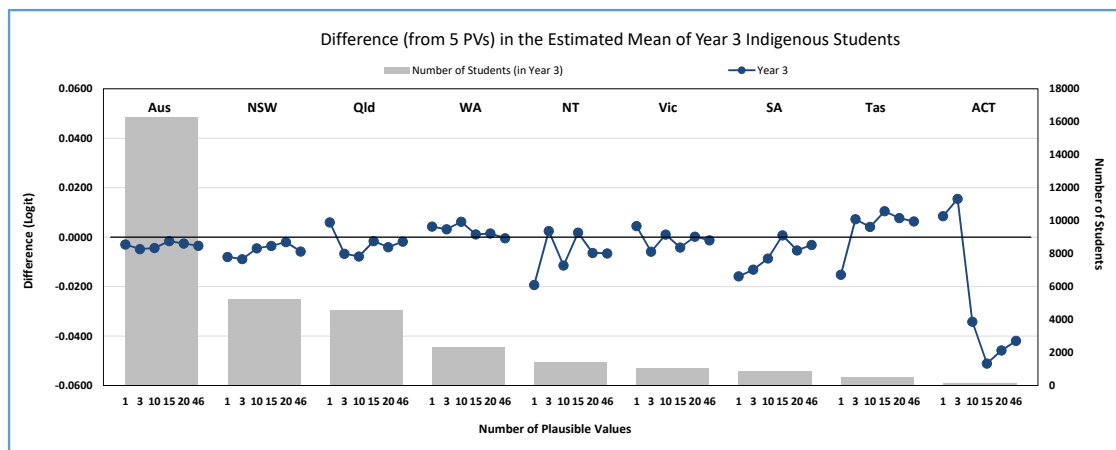
The parameter estimates and their standard errors in this section were based on the same plausible values that were used in the previous section. Again the results estimated from each of the 100 sets of plausible values were grouped into seven groups, each group aggregated from a different number of plausible values as described in the previous section. The effects on parameter estimates for overall Australian Indigenous students as well as by jurisdiction are discussed in the following sections. The detailed outcomes of the NAPLAN 2015 Numeracy tests by Indigenous students by the number of plausible values used are given in Appendix 7.

The Parameter Estimates

The differences between what was estimated using M plausible values and the reference value are calculated by year level, and by jurisdictions and for Australia overall. Figure 5.7 to Figure 5.11 present the differences in the parameter estimates for NAPLAN 2015 Indigenous students. Again there is one plot per year level per parameter. The plot is arranged in the order of sample size of Indigenous students in Australia and in eight jurisdictions. The arrangement of jurisdictions in the plots is different from the arrangement of the plots in the previous section as the order of sample sizes are different for Indigenous students. The number of Indigenous students varied from 151 (or less depending on the year level) in the Australia Capital Territory to about 4,483 (or more depending on the year level) in New South Wales, and about 15,000 students across the whole of Australia. Within Australia and each of the jurisdictions, the difference of the parameter estimates is arranged by the number of plausible values used, from 1 to 46. The difference of the parameter estimates are plotted as dot-line graphs and the number of Indigenous students as bars on the same graph.

Figure 5.7 to Figure 5.11 show that for jurisdictions with a relatively large sample size (Australia overall, New South Wales, Queensland and Western Australia), the differences in parameter estimates between those that were estimated using M plausible values and the reference value were within ± 0.04 logits for the mean, standard deviation, 20th and 80th percentile estimates, within $\pm 0.5\%$ for the estimated percentage above the top band.

When the sample size was less than 1,000 (the Australia Capital Territory, Tasmania and South Australia), the variations in the parameter estimates using different numbers of plausible values increased. For example, the differences in the percentages in the top band estimated for the Australia Capital Territory Indigenous students using M plausible values were as large as $\pm 2\%$. However, no increasing or decreasing trend in the parameter estimates was observed when increasing the number of plausible values used. The parameter estimates using one or three plausible values showed a relatively large difference to the reference value, while the differences between the parameter estimates using 10 or more plausible values and the reference value were much smaller. This indicates that when the number of plausible values used increased to 10 or more, the parameter estimates were close to the reference values. However, this was not always true when the sample size was extremely small, as in the sample of 151 or less for the Australia Capital Territory.



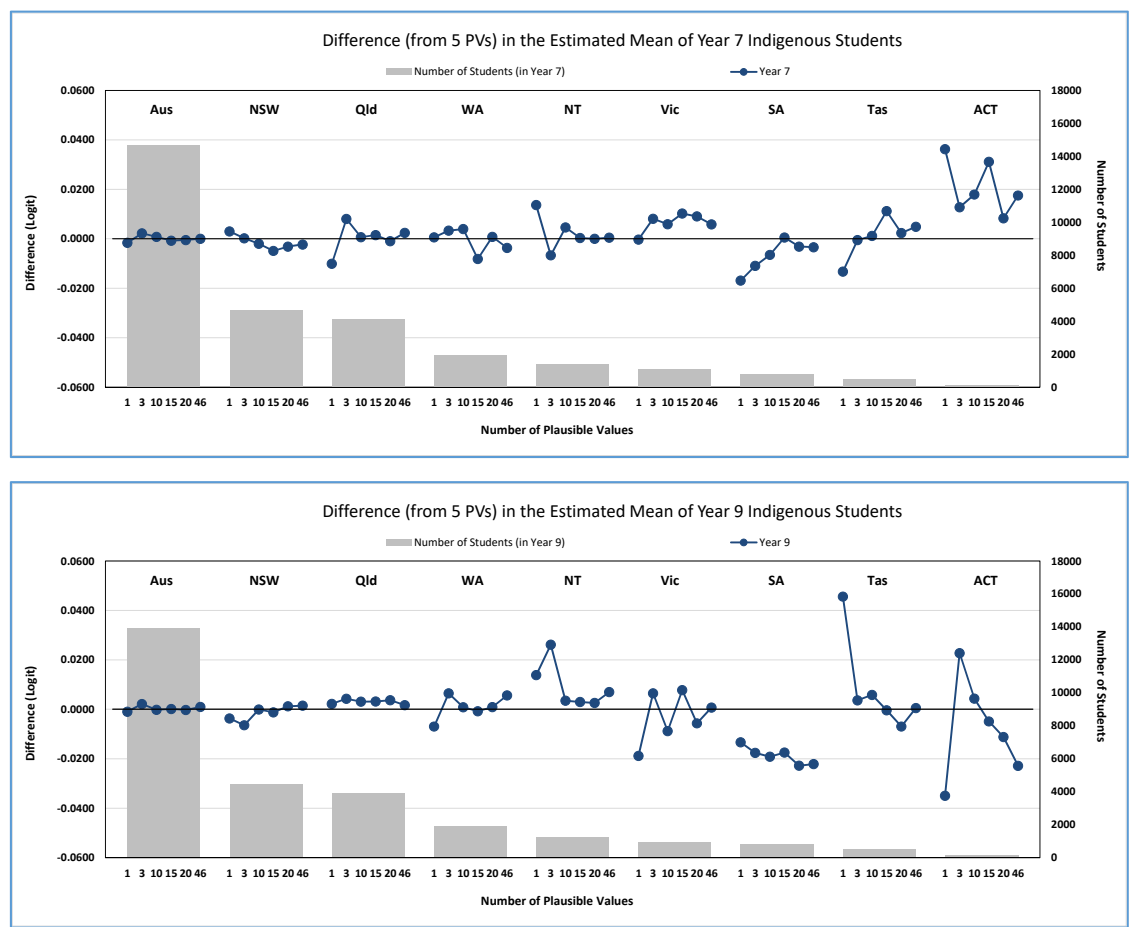


Figure 5.7 The Differences in the Mean Estimates in Empirical Study Two

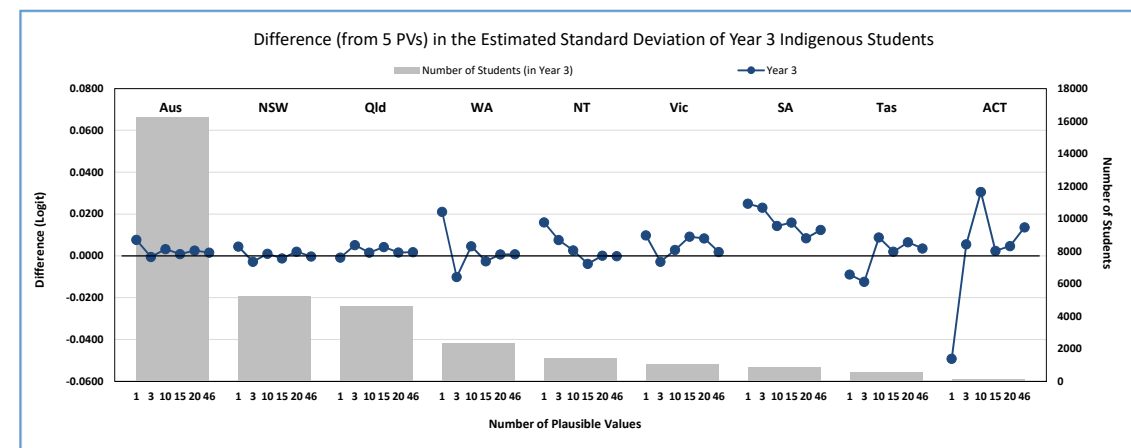




Figure 5.8 The Differences in the Variance Estimates in Empirical Study Two



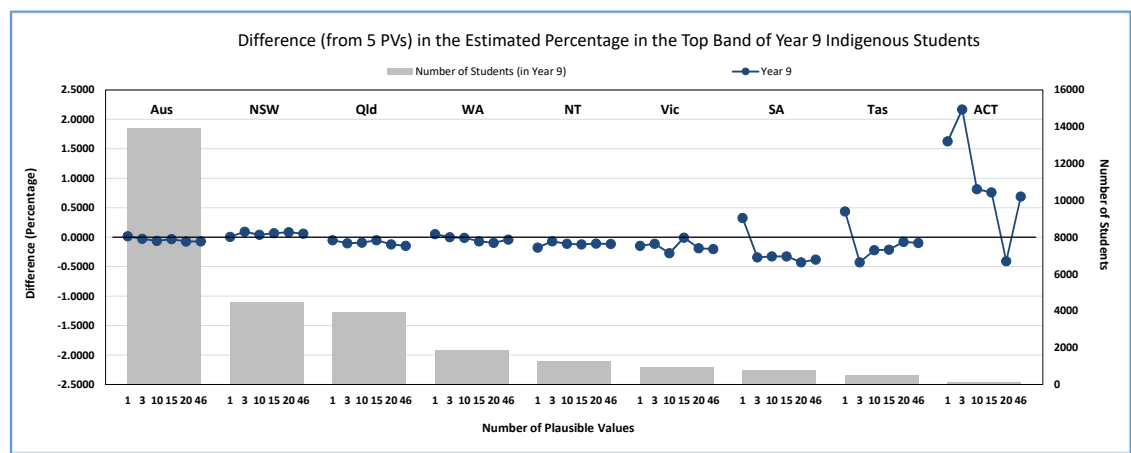
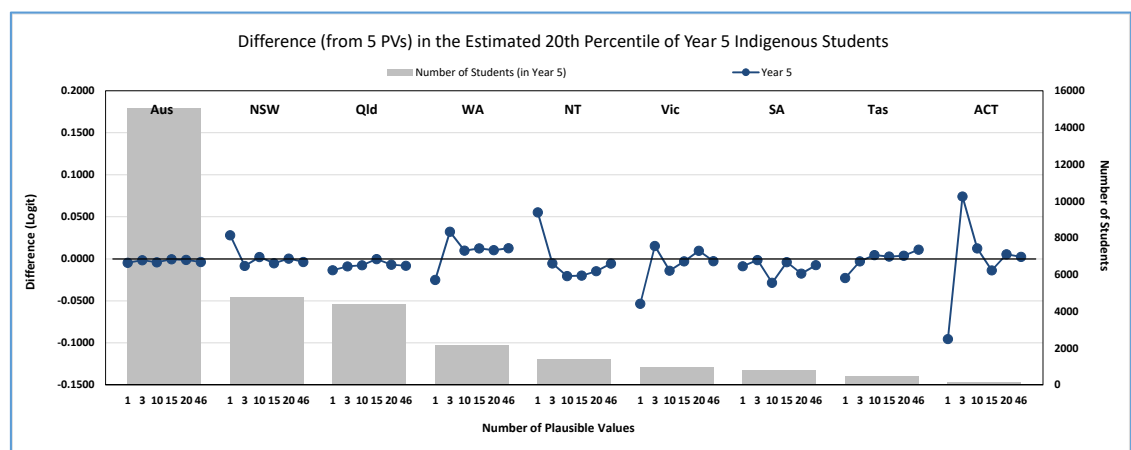
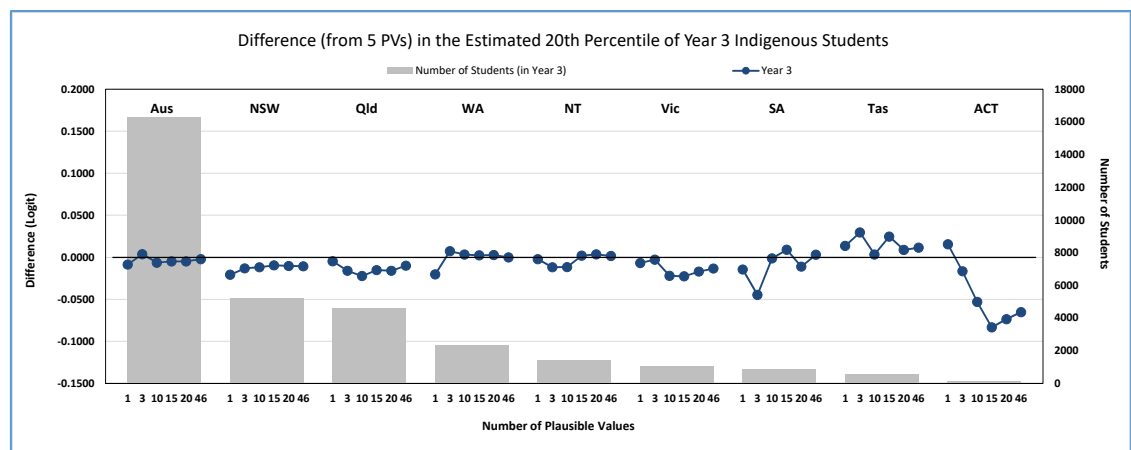


Figure 5.9 The Differences in the Percentage in the Top Band Estimates in Empirical Study Two



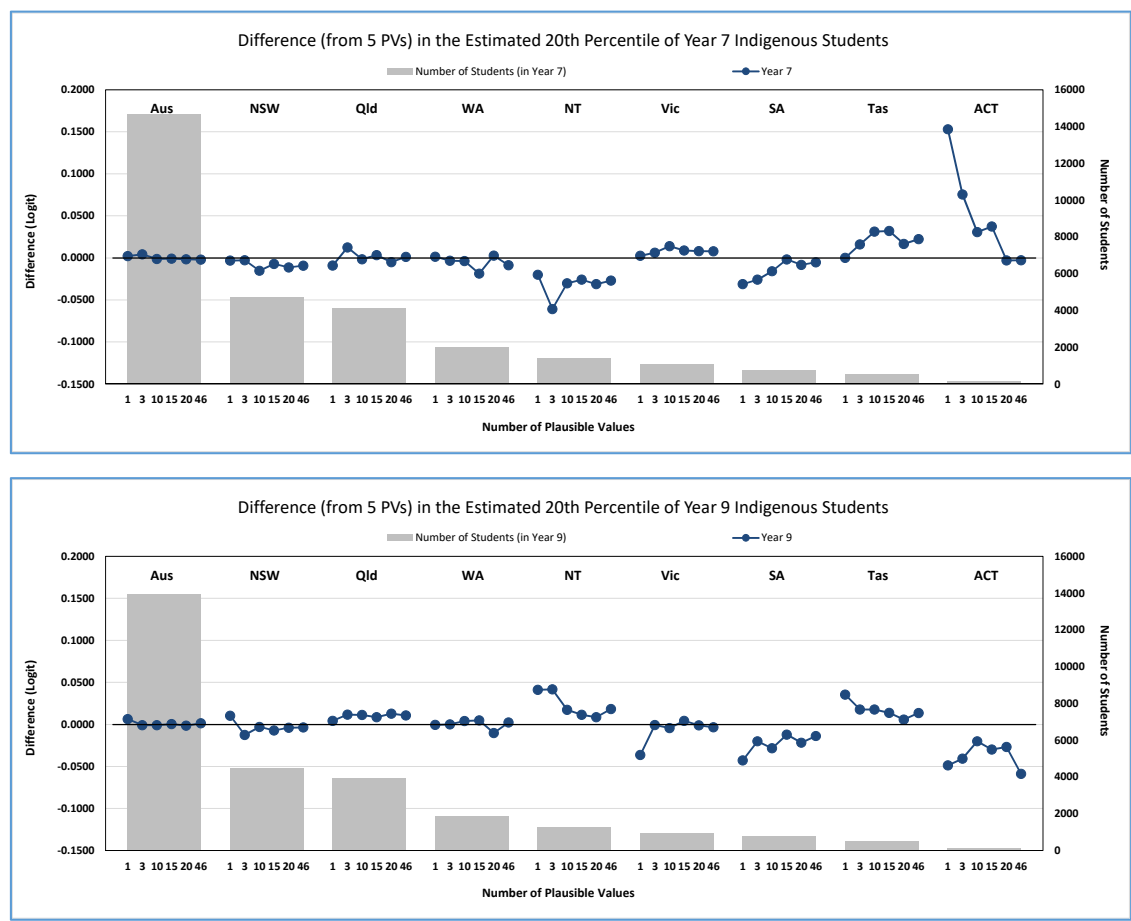
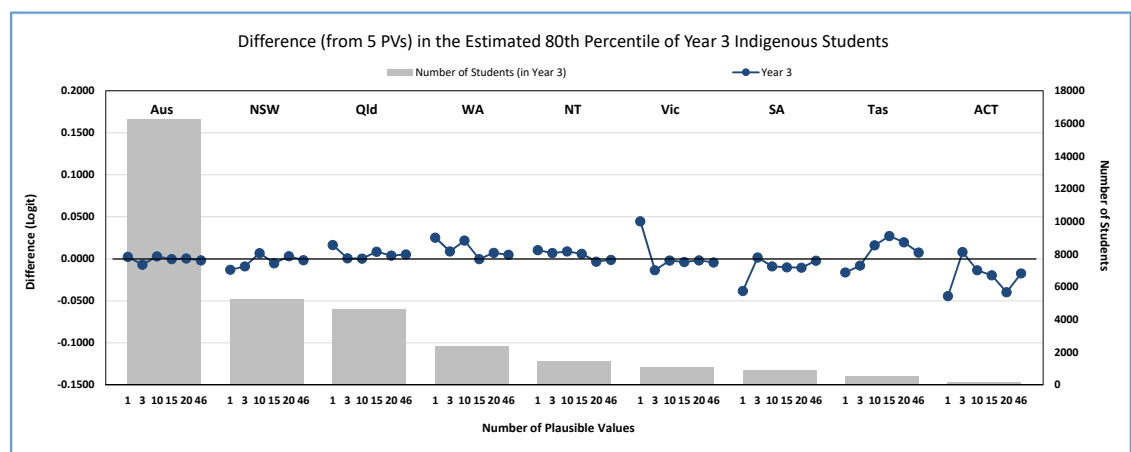


Figure 5.10 The Differences in the 20th Percentile Estimates in Empirical Study Two



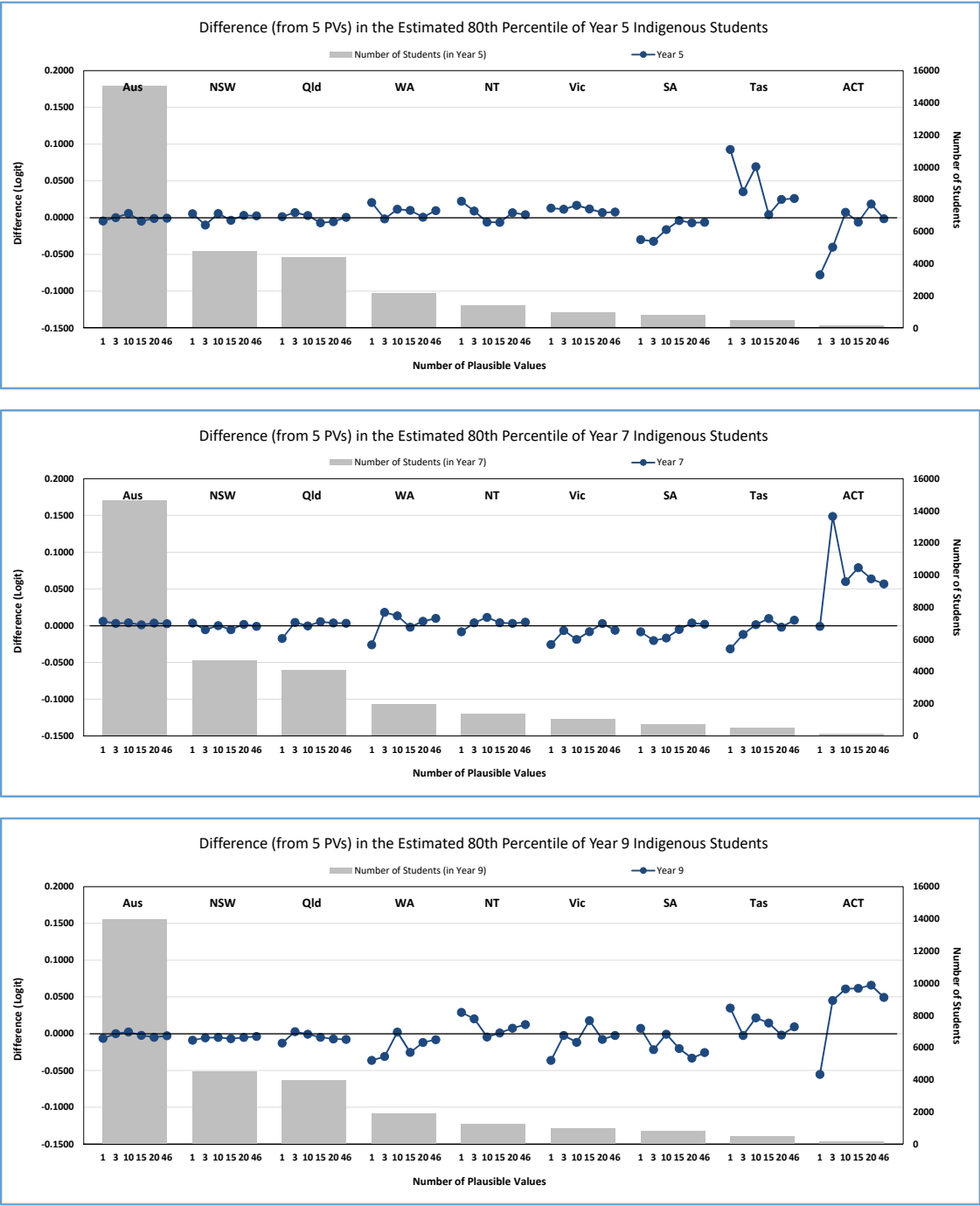
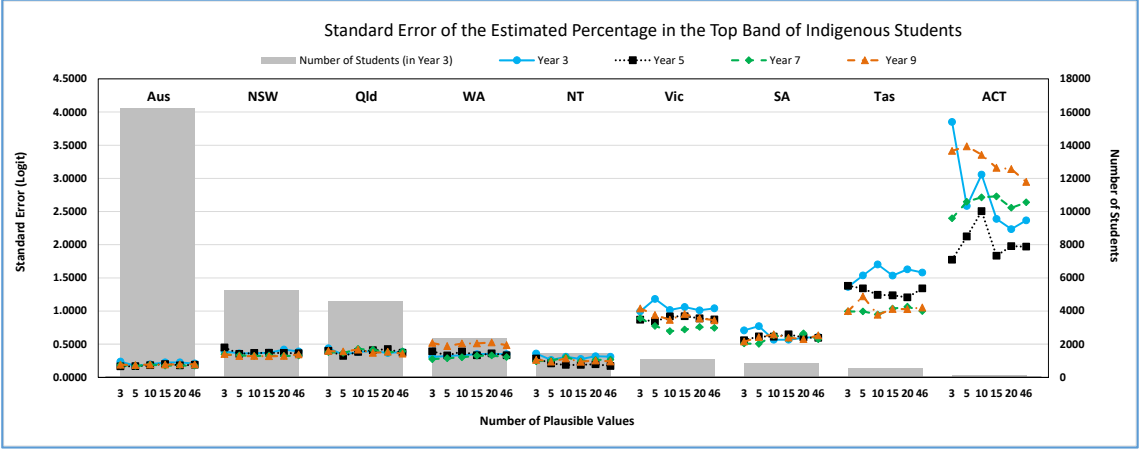
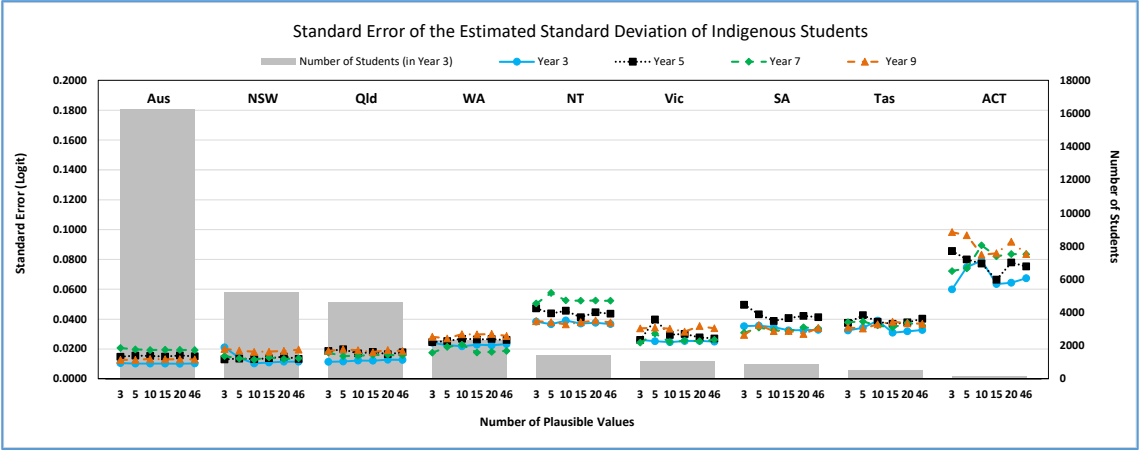
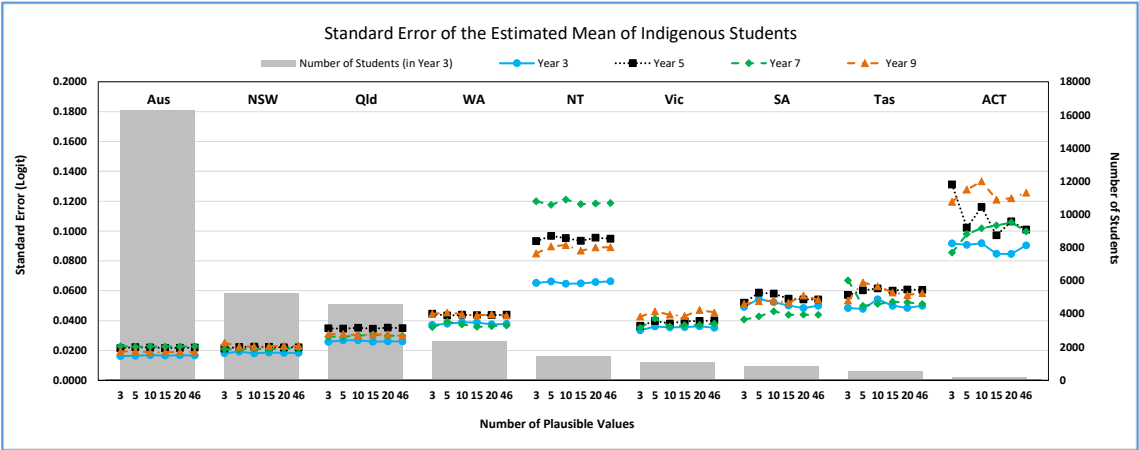


Figure 5.11 The Differences in the 80th Percentile Estimates in Empirical Study Two

The Estimated Standard Error of Parameters

Figure 5.12 presents the standard errors of the parameter estimates using the Indigenous student data. Similar to the standard error plots in the previous section, there is one graph for each standard error of parameter estimates. Each graph displays the standard errors of the parameter estimates by year level, as well as the number of Indigenous students in Australia overall and by jurisdiction. The estimated standard errors are arranged by the number of plausible values used (from 3 to 46) for Australia and for each of the jurisdictions. The numbers of Indigenous students in Year 5 for Australia and the jurisdictions are plotted in each graph in a bar chart, as an approximate indication of sample size. The standard errors are shown as dot-line graphs and the number of Indigenous students are shown as bars in each graph.

Similar to the standard error graphs in the previous section, Figure 5.12 shows that the standard errors of parameter estimates changed little with the different number of plausible values used. The difference between the smallest and the largest standard error by jurisdiction was less than 0.06 logits for the mean, standard deviation, 20th and 80th percentile estimates and less than 1.6% for the estimated percentage in the top band. There was no clear trend observed with an increase in the number of plausible values used in the estimated standard errors. In the Australia Capital Territory (less than 152 students), the variations among the estimated standard errors using the different number of plausible values were larger than the variations from other jurisdictions which had a larger sample size. The estimated standard errors generally increased with a reduction in the number of students. However, there was one exception observed in the standard error of the estimated percentage in the top band, the estimated standard errors in the Northern Territory were smaller than the estimated standard errors for New South Wales, Queensland and Western Australia which all had more Indigenous students than the Northern Territory. The small standard errors of the estimated percentage in the top band were due to the fact that there was less than one percent of Indigenous students in the Northern Territory (less than 10 students in each year level) who performed at or above the Australian national 80th percentile.



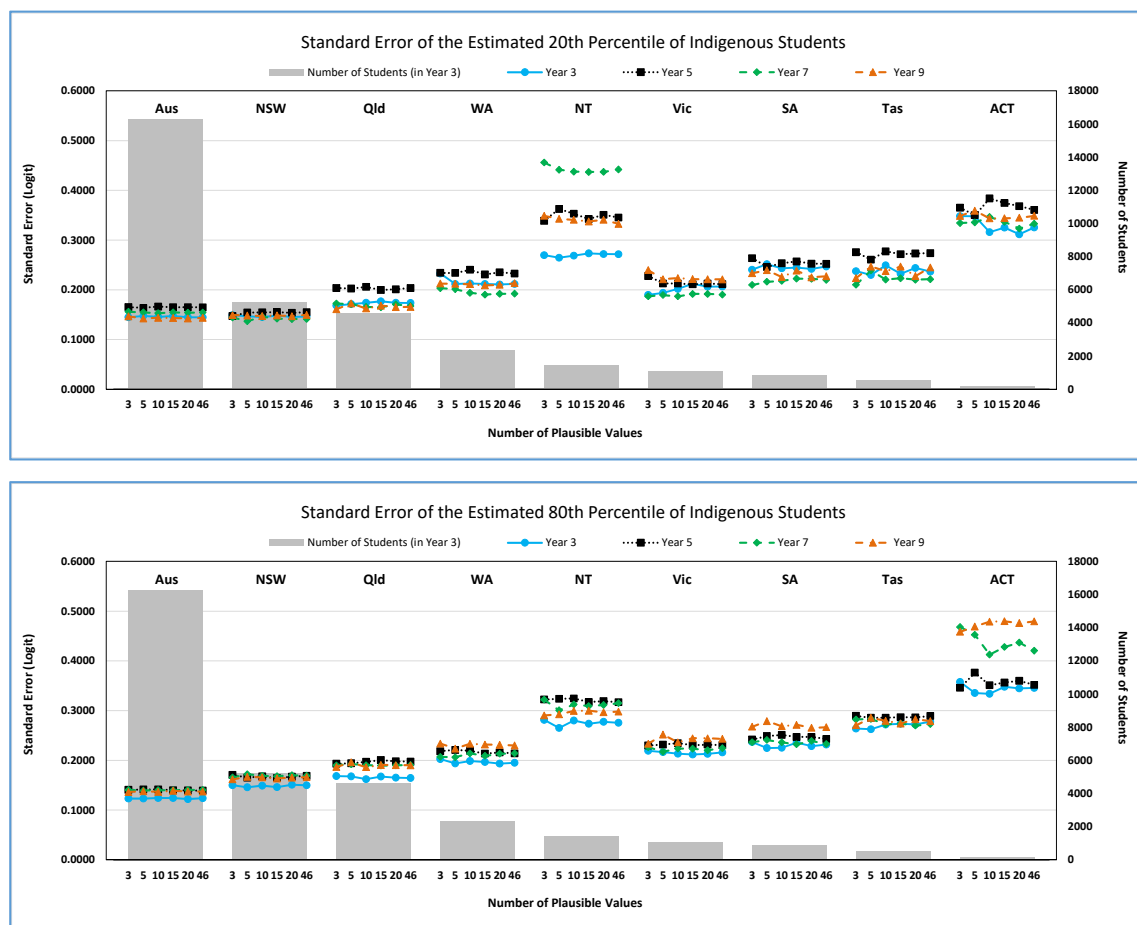


Figure 5.12 The Standard Errors of the Parameter Estimates in Empirical Study Two

In summary, the findings from analysing the NAPLAN 2015 data from Indigenous students were similar to the findings from the analysis of the NAPLAN 2015 whole cohort. There was no effect found in the parameter estimates or in the standard error estimates with increasing the number of plausible values from either of the two empirical data studies. The magnitude of the variation in parameter estimates were increased with the number of plausible values. In general, the smaller the sample size, the larger the estimated standard error.

Chapter 6 Discussion

This chapter summarises and discusses the findings from both simulations as well as the empirical data study. There are five research questions this dissertation intended to investigate. Section 6.1 provides the summary of findings from each study. Section 6.2 discusses the outcomes by research question. Section 6.3 discusses the impact of the bias and RMSEs of the simulation studies from a real practice point of view.

6.1 The Summary of Findings by Each Study

Chapter 4 reported the results from two simulation studies and Chapter 5 reported the results from two empirical data studies. The findings from these studies are summarised below.

Simulation one:

- There was no evidence that the number (in the range of 1 to 20) of plausible values used was related to the bias in the six parameters that were considered (mean, variance, percentage of students in the top band, 90th, 95th and 97.5th percentiles).
- The RMSEs of the six parameter estimates reduced with an increase in the number of plausible values used. When the number of plausible values increased from 1 to 20, with a large sample size ($N > 2,000$), the reduction in RMSEs were less than 0.0072 logits (or by less than 4% on average) across all simulated conditions compared with using only one plausible value; with small sample sizes ($N = 200$), the magnitude of reduction in RMSEs was less than 0.0539 logits or by less than 10% on average.
- The average of (using Rubin's method) the standard errors of the mean, variance, percentage in band, and percentile estimates reduced with an increase in the

number of plausible values used. The reduction does not improve the accuracy of the estimates of standard errors (see Section 6.2.2 for more details).

- With improved test targeting conditions, it was shown that the bias in the estimated means reduced, but there was little effect on the bias in the other parameter estimates. The only case where test targeting conditions showed an effect on the RMSEs of parameter estimates was the extremely off-target (2.5 logits) condition where the RMSEs were nearly twice the size of the corresponding values on the better-targeted tests. However, these large RMSEs increase did not coincide with an increase in the average estimated standard errors (see Section 4.1.2 for more details). That is, there was no effect observed on the estimated standard error when changing the test targeting conditions.
- The biases and the RMSEs for all six parameter estimates reduced with an increase in the test length, except for the sample from the normally distributed population, there was no consistent reduction pattern observed in estimated bias with an increase in test length.
- The estimated standard errors of the parameter estimates reduced with an increase in test length, with one exception, that was, the estimated standard errors increased with increasing test lengths for 95th and 97.5th percentiles when sampling from a chi-squared ($df = 5$) population distribution.
- When students were sampled from a normal distribution, the bias in all parameter estimates was smaller than the bias when students were sampled from non-normal distributions.
- The RMSEs of the mean and variance estimates did not show different patterns across the four population distributions. The 90th, 95th and 97.5th percentiles from the bimodal distribution had larger RMSEs than those from the other three distributions.
- Increasing sample size reduced the bias in the estimated mean across all simulation conditions with one exception, that is with an on-target test and cases sampled from a chi-squared ($df = 5$) distribution. With this simulated condition, there was little difference in the bias in the estimated mean across the three different sample sizes. There was a small reduction in the bias in the estimated

variance, but no reduction in the bias in the other parameter estimates. The RMSEs of the parameter estimates reduced with an increase in sample size.

Simulation two:

- There was no evidence that the number of plausible values used was related to the bias in parameter estimates as was found in Simulation One.
- It was observed that the RMSEs and the standard errors of the parameter estimates, including the regression coefficient estimates, reduced with an increase in the number of plausible values used.
- There was no effect observed on the bias in the parameter estimates with an increase in the test length. However, it was observed that the RMSEs and the standard errors of the parameter estimates reduced with an increase in test length when the test length is 40 items or less. Further, after increasing the test length to 60 items, the reductions in the RMSEs and the standard errors of parameter estimates was very small if any.
- There was no evidence that test target conditions had any effect on the recovery of parameters, in terms of the bias, the RMSEs and the standard errors.

Empirical Study One:

- The differences in the parameter estimates from using one plausible value to using 46 plausible values were less than ± 0.01 logits for the mean, standard deviation and the 20th and 80th percentile estimates, and were less than 0.15% for the percentage of students in the top band for jurisdictions with a large number of students ($N > 6,000$).
- The variations in the parameter estimates using different numbers of plausible values increased when the number of students decreased, but there was no trend observed in the parameter estimates with an increase in the number of plausible values used.
- The parameters estimated using five or more plausible values were less variable than the parameters estimated using fewer than five plausible values.

- Using different numbers of plausible values (from 3 to 46), the differences between the standard errors of a parameter estimate was less than 0.02 logits for the mean, standard deviation and the 20th and 80th percentile estimates, and was less than 0.16% for the estimated percentage of students in the top band.
- There was no evidence that test targeting status had an effect on parameter estimates or their standard error estimates.
- Unlike the findings from the simulation studies, there was no decreasing trend (and no increasing trend either) observed in the standard errors of the parameter estimates with an increase in the number of plausible values used. This implies that increasing the number of plausible values used in estimating standard errors from 3 to 46 had a limited effect on the standard error of estimates for the NAPLAN 2015 data.

Empirical Study Two:

- It was observed that for New South Wales, Queensland and Western Australia, and for Australia overall, all of which have sample sizes of more than 5,000 students, the differences in the parameter estimates from using 1 to 46 plausible values, were less than ± 0.04 logits for the mean, standard deviation, 20th and 80th percentile estimates, and were less than 0.5% for the estimated percentage of students in the top band.
- A relatively large variation in the standard error of parameter estimates from a very small sample size ($N \leq 151$) using different numbers of plausible values was observed. This implies that increasing the number of plausible values used may not result in better standard error estimates.

6.2 Discussion by Research Questions

6.2.1 The Effect of Increasing Numbers of Plausible Values Used on the Population Parameter Estimates

The first research question of interest in this dissertation is: How would varying the number of plausible values used affect the population parameter estimates?

There was no evidence that the number of plausible values used (ranging from 1 to 20) was related to the bias in the parameter estimates (mean, variance, percentage of students in the top band, 90th, 95th and 97.5th percentiles).

As summarised in the previous section, there was no evidence that the bias in the parameter estimates was related to the number of plausible values used in any of the four studies. From the NAPLAN 2015 empirical data studies, there was no effect observed on the RMSEs of population parameter estimates with an increase in the number of plausible values used regardless of the sample size. From the two simulation studies, however, a reducing trend was observed in the RMSEs of the parameter estimates (including regression coefficients) with an increase in the number of plausible values used, especially when the sample size was small ($N = 200$).

6.2.2 The Effect of Increasing Numbers of Plausible Values Used on the Standard Error Estimates

The second research question of interest in this dissertation is: Will the size of the standard error of parameter estimates be reduced by increasing the number of plausible values used?

The standard errors of all parameter estimates reduced with an increase in the number of plausible values used across all simulation conditions. This reduction trend, however, was not observed in the NAPLAN 2015 empirical data studies.

The total standard error of a parameter estimate is estimated by combining the sampling variance ($\bar{U}$) and the measurement variance (B_M). The formula to combine the two variances was defined by Rubin (1987) as shown in Equation 3.10.

$$V = \bar{U} + (1 + M^{-1})B_M \quad \text{Equation 6.1}$$

In this formula, there is a factor $(1 + 1/M)$ applied to the measurement variance which is an adjustment for the finite number of plausible values used (Rubin, 1987; Schafer and Olsen, 1998). It is obvious that with an increase in the number (M) of plausible values used, the adjustment factor became smaller (Figure 6.1). This means that the measurement variance component in Equation 6.1, after applying the adjustment factor, will reduce with an increase in the number of plausible values used for the same B_M . The adjustment factor will approach to one with an increase in M . It shows that from Figure 6.1 the adjustment factor is approaching to one, once M reaches to 20, the adjustment factor would have little effect in Equation 6.1.

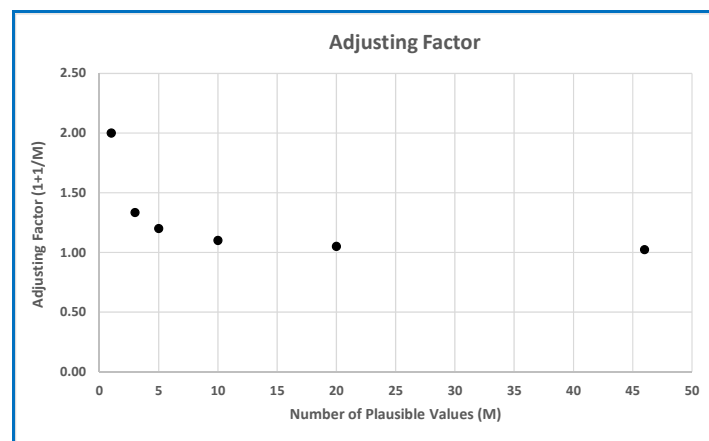


Figure 6.1 The Adjustment Factor by Number of Plausible Values

Although the sampling error may change slightly across each set of plausible values for the same sample, the variation is generally negligible. In fact, the sampling errors were estimated with only one set of plausible values regardless of the number of plausible values used for parameter estimates in the NAEP reporting procedure (Beaton 1987). The

variance among plausible values would be approximately the same across a different number of plausible values as long as the objective to be measured is the same, and the same measurement model is applied. Thus, according to Rubin (Equation 6.1) (Rubin, 1987) the combined total of standard errors would reduce with an increase in the number of plausible values used.

On the other hand, if we assume:

$$\bar{U} = kB_M \quad \text{Equation 6.2}$$

Where k is the ratio between the sampling variance and measurement error variance for a set of data then substituting this into the Equation 6.1 it becomes:

$$V = (1 + M^{-1} + k)B_M \quad \text{Equation 6.3}$$

Equation 6.3 reveals that with the same ratio k , increasing the number of plausible values used will reduce the total variance. On the other hand, with the same M , increasing the ratio k between the sampling variance and the measurement variance will increase the total variance, therefore, increasing the standard error.

Figure 6.2 depicts the effect of adjustment factor $(1 + 1/M)$ on the total combined standard errors. It also shows the standard errors of the means from 2,000 students sampled from a normally distributed $N(0, 1)$. The graph is arranged by four test lengths from 10 items to 60 items, then by the four test targeting statuses from an on-target to an extreme off-target test. Three sets of variances are plotted: the green triangles represent the sampling variances, the red diamonds represent the variances among plausible values and the blue dots represent the combined total variances according to Equation 6.1.

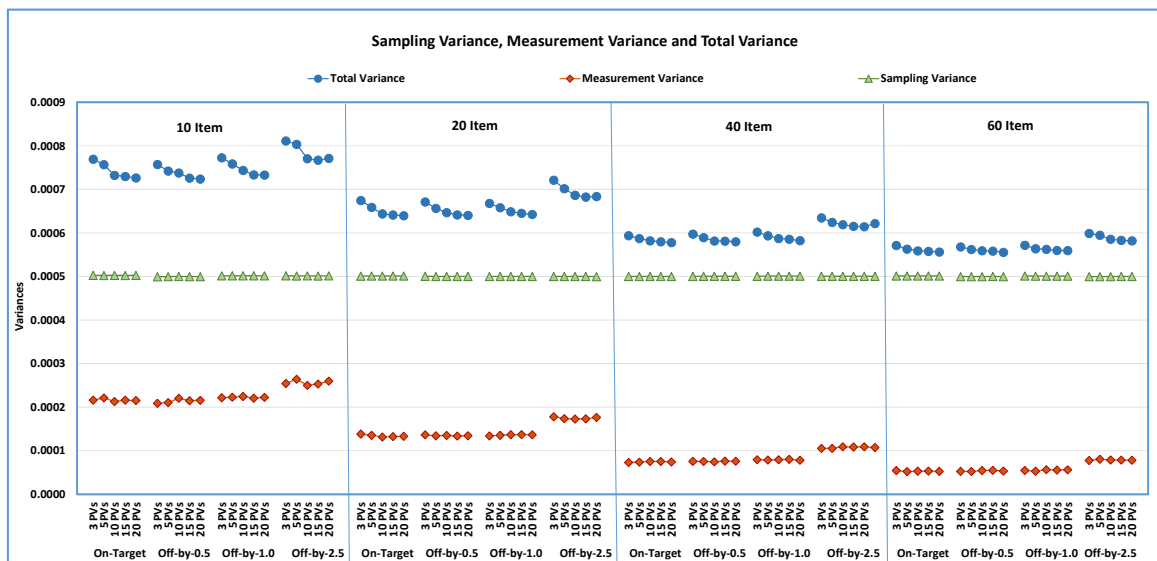


Figure 6.2 The Sampling, Measurement and Total Variances of the Mean Estimates from Simulation

Figure 6.2 shows that the sampling variances were approximately the same across four different test lengths, for the four test targeting conditions, and for the same sampling condition (2,000 cases from a normally distributed population). The measurement error variances were approximately the same within the same test conditions across different numbers of plausible values. They reduced with an increase in the test length, and increased when the targeting condition became extremely off-target, which is due to the measurement variance increasing when the test reliability increases. This shows that the combined total variance was reduced with an increase in the number of plausible values used within each test condition. Both measurement variance and sampling variance remain approximately unchanged using different numbers of plausible values. The reduction in the total variance estimates is mainly due to the adjustment factor $(1 + 1/M)$, and not due to either the reduction in the variation among the M plausible values used, nor to the reduction of sampling variance.

In the NAPLAN 2015 empirical data study, a reduction in the standard errors of the parameter estimates was not observed with an increase in the number of plausible values used. When examining the sampling variances and the measurement variances of the empirical data separately, as depicted in Figure 6.3, it shows that the total variances were

mostly dependent on the sampling variances. In addition, the measurement variances fluctuated considerably when the sample size was very small ($N \leq 151$).

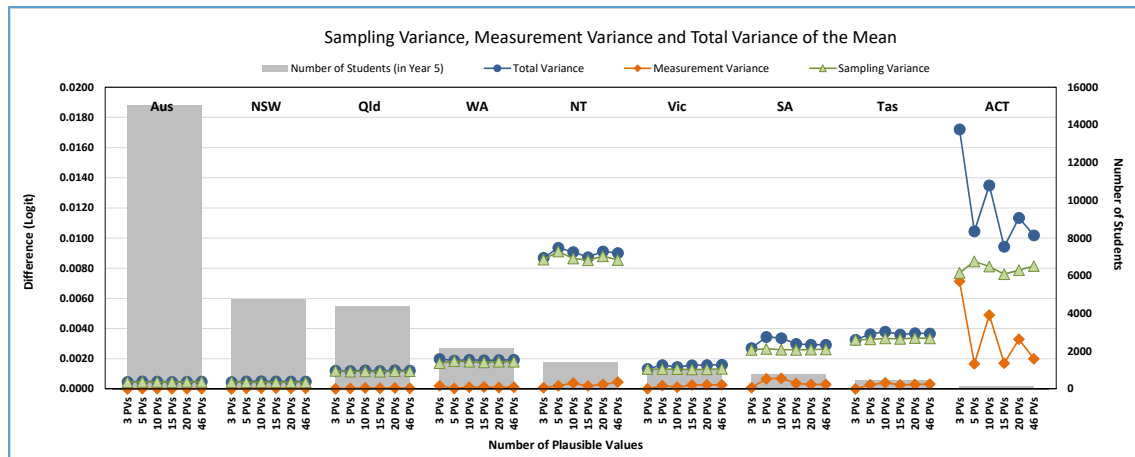


Figure 6.3 The Sampling, Measurement and Total Variances of the Means from the NAPLAN Data

Table 6.1 provides the proportion of variance: due to sampling and due to the measurement associated with the latent nature of student ability estimates. There are three sets of results: the first simulation study, NAPLAN 2015 WA Indigenous students and the NAEP 1996 Grade 4 Reading test (Allen, Johnson, Mislevy & Thomas, 1999). This particular simulation and the NAPLAN data were selected because of their similar condition in terms of sample size, test length and test targeting status.

Table 6.1 shows that the sampling variance was a much larger proportion ($> 80\%$) of the total variance than the measurement variance for all three sets of results. Further, the proportion of the sampling variance in the simulation data is about 85% of the total variance, which was lower than the proportion of the sampling variance in the empirical data: about 93% in NAPLAN and about 92% in NAEP (Allen *et al.* 1999). The larger sampling error in the empirical data is due to its hierarchical data structure which resulted in a much smaller effective sample size, hence a larger sampling variance. The ratio between the sampling variance and the measurement variance from the simulation data remained about the same, at 6.6% to 6.7% using different numbers of plausible values. This stability was not observed in the NAPLAN data.

There are two factors in Equation 6.3: the adjustment factor for the number of plausible values ($1/M$) and the ratio (k) between the sampling variance and the measurement variance. For fixed k , total variance is dependent on the number of plausible values used, and reduces with an increase in the number of plausible values used. When the ratio (k) varied considerably, the total variance will depend on both factors. When the ratio (k) between the sampling variance and the measurement variance was much larger than the adjustment factor ($1/M$), the effect of the number of plausible values used diminished. For this reason, there was no reducing trend in the standard errors with an increase in the numbers of plausible values used observed in the NAPLAN 2015 data.

Table 6.1 Estimated Error Variance and Related Parameters

Data Source	Number of PVs (M)	Sampling Variance: ($\bar{U}$)	Measurement Variance: ($(1+1/M)B_M$)	Total Variance: (V)	Proportion of Students Sampling: ($\bar{U} / V$)	Variance Due to Latency of (θ) ($(1+1/M)B_M/V$)	Ratio of $\bar{U} / B_M$ (k)
Simulation One (Normally distributed sample 2000, 40 items off-target by -1.0 test)	3	0.00050	0.00010	0.00060	0.83	0.17	6.6
	5	0.00050	0.00009	0.00059	0.85	0.15	6.6
	10	0.00050	0.00008	0.00058	0.86	0.14	6.7
	15	0.00050	0.00008	0.00058	0.86	0.14	6.6
	20	0.00050	0.00008	0.00058	0.86	0.14	6.6
NAPLAN Year 5 WA Indigenous Students (2,130 students, with Target = -1.090 from a 40 items test)*	3	0.00171	0.00028	0.00199	0.86	0.14	8.2
	5	0.00185	0.00004	0.00189	0.98	0.02	59.5
	10	0.00180	0.00013	0.00193	0.93	0.07	15.2
	15	0.00177	0.00014	0.00191	0.93	0.07	13.7
	20	0.00180	0.00011	0.00191	0.94	0.06	17.4
	46	0.00181	0.00012	0.00193	0.94	0.06	15.0
NAEP 1992 Grade 4 Reading (Allen et al. 1999, p. 250) **	5	0.833	0.076	0.908	0.92	0.08	13.2*

* Estimated value from the available data

** In NAEP scale scores

Although the results of the simulation studies showed that the estimated standard errors reduced with an increase in the number plausible values used, this reduction in the estimated standard errors may not be desirable as demonstrated below:

The standard error of the mean could be determined using three different methods. These methods are: the estimation method proposed by Rubin (1987) and described in Chapter 3; (square root of) the observed variance among replications of the simulation; and an analytic method. The following sections provide an example of standard errors that were estimated with each of the three methods. The standard errors estimated by the observed method and the analytic method are estimated with one (the first set) plausible value only. The estimation of standard errors by the analytic method used the following formula:

$$SE = \sqrt{\frac{\sigma^2}{N\rho}} \quad \text{Equation 6.4}$$

where: σ^2 is the sample variance, N is sample size and ρ is the average test reliabilities over 1,000 replications.

It is known that the reliability of a test is a function of test length and test targeting (Lord, 1959). Although the item parameters were fixed on the generating values for the simulation studies in this dissertation, student response data were randomly generated based on the simple logistic model and the selected simulation conditions. Therefore, the reliability alters slightly for each replication of the simulation. For simplicity, the average of the reliabilities from the generated data was used.

The standard errors of the estimated mean were estimated using these three methods by sample size, by test targeting status and by test length. The detailed results are given in Appendix 8 and presented in Figure 6.4. The same colour scheme and symbols used in the figures in Chapter 4 are used in Figure 6.4. To illustrate, Table 6.2 presents the standard errors estimated using the three methods for a normally distributed sample with 2,000 cases, and a 40-item on-target test as an example.

Table 6.2 Comparing the Standard Errors of the Mean When Estimated by Three Methods Using a Normally Distributed Sample ($N = 2,000$) with On-Target Tests

Test Length	SD among replicates	Analytic method	Rubin's Method with Different Number of PVs				
			3PVs	5PVs	10PVs	15PVs	20PVs
10	0.0338	0.0295	0.0281	0.0277	0.0271	0.0271	0.0270
20	0.0290	0.0261	0.0262	0.0258	0.0254	0.0253	0.0253
40	0.0265	0.0243	0.0244	0.0243	0.0241	0.0241	0.0240
60	0.0245	0.0237	0.0240	0.0237	0.0237	0.0236	0.0236
80	0.0236	0.0234	0.0235	0.0234	0.0234	0.0233	0.0233

Figure 6.4 shows that, for the three estimation methods, the results from the observed standard errors method (labelled as “SD Among replicates”) were the largest; with Rubin’s method resulting in the smallest estimated standard errors when the test was short (less than or equal to 20 items), especially when sample sizes were small ($N = 200$) or medium ($N = 2,000$), or the test was extremely off-targeted (2.5 logits). The observed standard error can be seen as the true error. It shows that the analytic method and Rubin’s method actually under estimated the total variance, and increasing the number of plausible values used did not make the estimated total variance closer to the observed variance. This also means that the practice used in PISA 2015 of increasing the number of plausible values to “get better estimates of the variability” of parameter estimates is questionable.

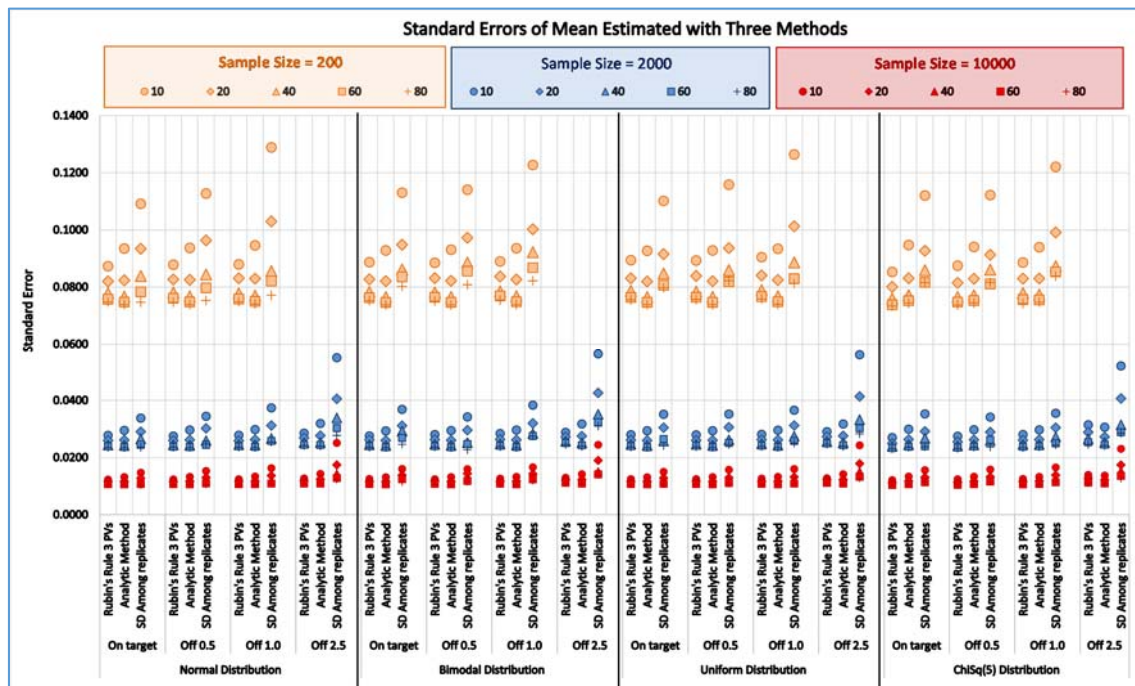


Figure 6.4 The Standard Errors of the Mean Estimated with Three Methods

The standard errors estimated from the observed variance among replications can be deemed to be the expected standard error. Comparing the bias and the RMSEs estimated using three plausible values in the previous sections with the expected standard errors, it was found that the ratios of absolute bias to the expected standard error were less than 0.05 for samples from a normal distribution when the sample size was 2,000 or more, and when tests were reasonably targeted (maximum off-target by 1 logit). The ratios of RMSEs to the expected standard errors were related to the sample size. The ratio was about 0.01 for the large samples ($N = 10,000$), up to 0.03 for the medium samples ($N = 2,000$) and up to 0.10 for the small sample ($N = 200$).

6.2.3 The Effect of Test Reliability When Using Plausible Values

The third research question of interest in this dissertation is:

How does the IRM reliability affect the population parameter estimates when estimated using plausible values, in terms of test length (number of items) and test targeting conditions?

The estimated ability for individual i with ability θ can be expressed by $\hat{E}[\theta|X = x_i]$, where x_i is item responses. The measurement variance of the estimated ability is the reciprocal of Fisher Information $I(\theta_i)$:

$$variance(\hat{E}) \propto \left(\frac{1}{I(\theta_i)} \right) \quad \text{Equation 6.5}$$

The estimate ($\hat{t}$) of a parameter of interest (t) from a sample (sample size = N) can be expressed as:

$$\hat{t} = \frac{1}{N} \sum_i \hat{E}[\theta|x_i] \quad \text{Equation 6.6}$$

Therefore, the measurement variability of the estimate ($\hat{t}$) of the interest parameter (t) can be approximated by:

$$variance(\hat{t}) = var\left(\frac{1}{N} \sum_i \hat{E}[\theta|x_i]\right) \propto \frac{1}{N} \sum_i \frac{1}{I(\theta_i)} \quad \text{Equation 6.7}$$

Equation 6.7 shows that when Fisher information $I(\theta_i)$ increased, the measurement variability of the estimator will be reduced, the reliability of the test will be increased.

Two factors that could affect the information, therefore affect the test reliability were investigated in the simulation studies: test length and test targeting. With the NAPLAN empirical data study, although there were different test lengths across different year levels, unfortunately these were confounded with the sample size variation and therefore no conclusions could be drawn regarding test lengths.

It is expected that with an increase in test length a test provides more information and becomes more reliable and, hence, the measurement error of the test becomes smaller. Thus, fewer plausible values are thought to be required for a long test compared with a short test to provide a similar accuracy of estimated parameters. Simulation One shows that the bias in the parameter estimates reduced with increasing test length, but with certain exceptions. When the parameters were estimated from a sample of normal distribution, there was no clear pattern shown in the bias in the parameter estimates with increasing test length. A pattern of reduced the bias in the parameter estimates with an increase in test length was not observed in simulation two. The RMSEs of parameter estimates reduced with an increase in test length, this pattern was also observed in simulation two and in the empirical data studies.

Standard errors of the maximum likelihood estimates are estimated by the inverse of the information matrix (Adams, 1997a). Compared to a short test, a longer test provides a more peaked likelihood, more test information and therefore yields parameter estimates with smaller standard errors. It follows that a longer test should provide more accurate parameter estimates and estimated standard error. This is demonstrated in Figure 6.5 which shows three likelihood curves plotted from tests with 20, 100 and 500 items, respectively. The three tests have the item difficulties spread in the same range (-3 to 3 logits) with the average item parameters set to zero. The likelihood curves were derived from responses based on the same student abilities which have responded 50% items correctly. The graph clearly shows that for a longer test, its likelihood is more peaked than the likelihood of a shorter test.

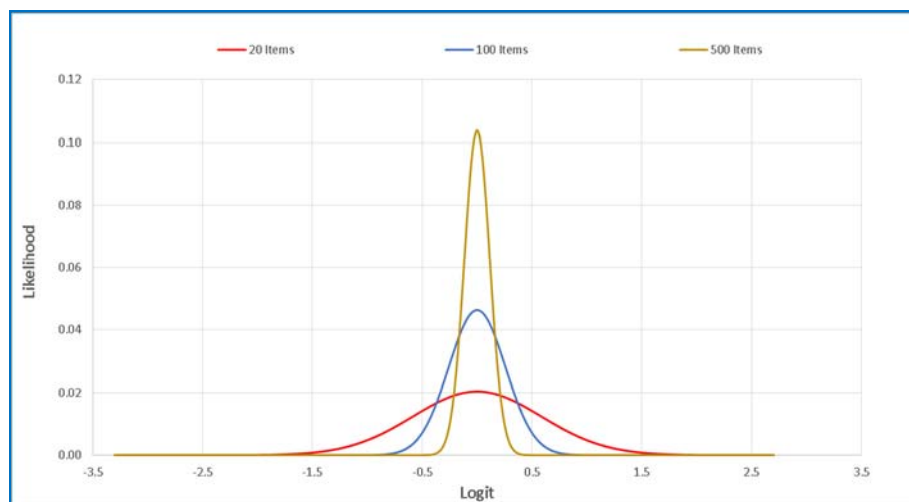


Figure 6.5 The Likelihood Functions with Test Length Increases

The test information function and standard error from the three different lengths of tests are depicted in Figure 6.6. This shows that the reliability of student ability estimates from a longer test was higher than the reliability of student ability estimates from a shorter test. This is reflected in the standard error of the proficiency estimates. Typically, with a 20-item test, the standard error of ability estimates is about 0.51, while the standard error of ability estimate based on a 100-item test is about 0.25, and the standard error of ability estimates based on a 500-item test is about 0.12 around the maximum information region.

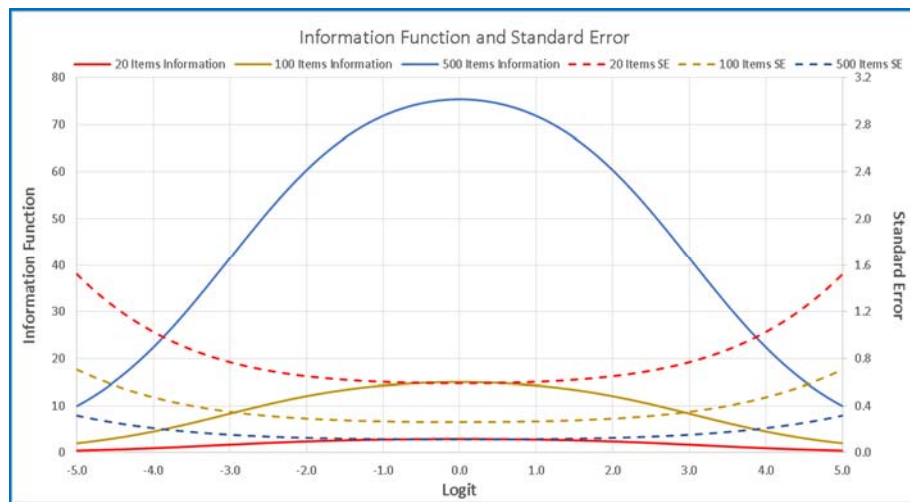


Figure 6.6 Test Information Function and Estimated Standard Error with Three Test Lengths

In terms of test targeting conditions, no relationship was observed between the test targeting condition and the bias and RMSEs of all the parameter estimates from the simulation studies, with a couple of exceptions. In Simulation One it was observed that the bias in the estimated mean increased with the deterioration of the test targeting status and; when tests were extremely off-target the RMSEs of the parameter estimates from the chi-squared ($df = 5$) distribution were larger than the RMSEs of the parameter estimates for the symmetrically distributed samples.

A simulation could not be carried out for a small sample with the extreme off-target condition because the extreme conditions resulted in perfect items during data generation (see explanation in Section 3.2). This resulted in the simulation with a small sample only covering a narrow range of test targeting conditions (0 to 1.0 logit off-target), which may prevent some effects of the number of plausible values used on the parameter estimates from being revealed. However, large-scale assessments generally have much larger samples than 200 students, and assessments should aim to match the targeted student ability range. In general, caution must be taken when drawing any conclusions from an extremely off-target assessment.

6.2.4 The Effect of Prior Distribution Assumption

The fourth research question of interest in this dissertation is:

How would the assumption of a prior distribution affect the parameter estimates?

This research question was investigated with simulation one which was carried out with samples drawn from four different population distributions. It was observed that the samples from normally distributed populations had the smallest bias in all the parameter estimates and no effect was shown in the RMSEs of the mean and variance estimates. The results from the bimodal distribution showed the largest RMSEs of the percentile estimates, with the results from the uniform distribution showing the largest RMSEs of the estimated percentage of students in the top band. In terms of the estimated standard errors, there was no effect observed except for the standard error of high percentile ranks in the skewed chi-squared ($df = 5$) distribution, e.g. 95th and 97.5th percentile rank.

A normal distribution, or conditional normal distribution with multiple regressions is assumed for population distribution in some IRT estimation software, for example, BILOG (Mislevy and Bock 1983), DGROUP (Rogers, Tang, Lin, & Kandathil, 2006) and ACER ConQuest4 (Adams *et al.* 2015). In practice, the sample distribution may deviate from a normal distribution. It has been found that with increasing test lengths, plausible values would be less affected by the assumed prior distribution (Marsman, Maris, Bechger & Glas, 2016). The shorter the test, the more the prior distribution will affect trait-level estimates. In long tests, on the other hand, the prior distribution will be overwhelmed by the likelihood function and will have little if any effect on trait-level estimation (Embretson and Reise, 2000). This suggests that parameter estimates of a long test require fewer plausible values compared with a shorter test. It is evidenced from Simulation One that the population parameter estimates were less biased and had smaller RMSEs with increased test length within each simulation condition, across all four population distributions.

In order to demonstrate the effect of test lengths on recovering the generating distribution, a small simulation of 2,000 cases sampled from a chi-squared ($df = 5$) distribution was carried out. First, a sample of 2,000 cases was drawn from a chi-squared ($df = 5$) population distribution (Figure 6.7(a)); and then it was standardised (Figure 6.7 (b)).

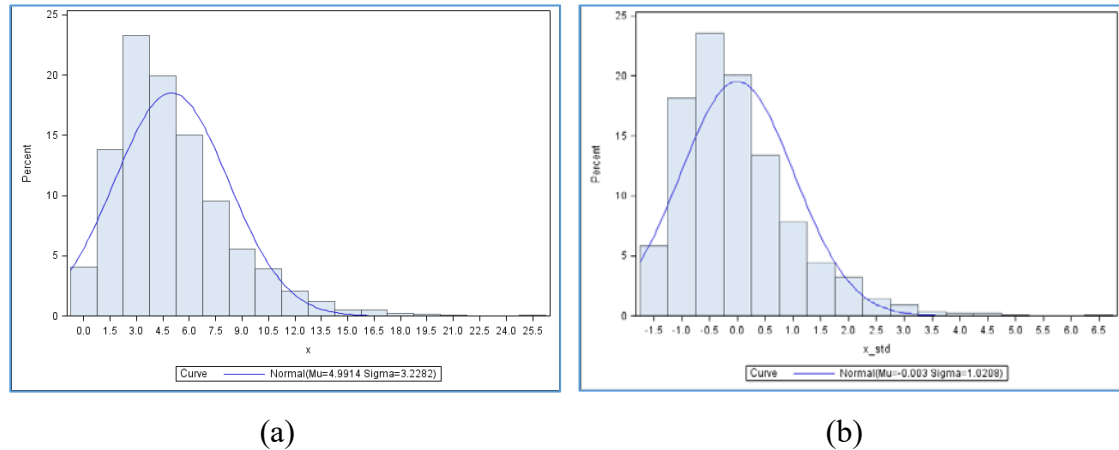


Figure 6.7 Chi-squared ($df = 5$) Distribution

A random sample of 100 cases (shown as pink in Figure 6.8) were selected from the 2,000 cases (shown as light blue in Figure 6.8), where θ_h represents the ability of student N ($N = 1$ to 100) (see Figure 6.8).

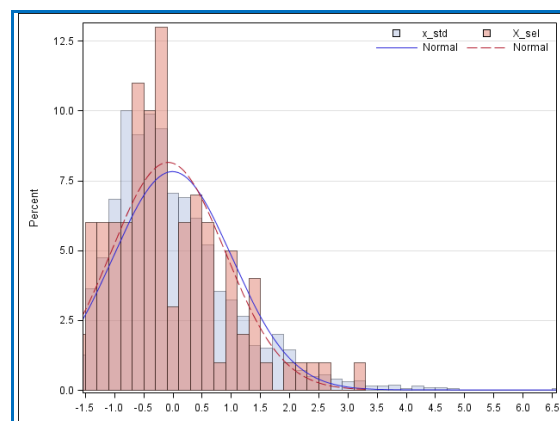


Figure 6.8 The Distribution of a Random Sample from the Original Chi-squared ($df = 5$) Distribution

In order to reflect the assumption that plausible values are drawn from a normally distributed ability range around the maximum likelihood, two sets of random samples of 200 were generated from a normally distributed population for each of these 100 students. The two samples have a mean equal to the student ability θ_n and standard deviations of 0.51 and 0.12, respectively. The sample with the large standard deviation of 0.51 depicts the posterior distributions from a shorter 20-item test. The sample with a small standard deviation of 0.12 depicts the posterior distribution from a much longer 500-item test (refer to Figure 6.6). Each observation within a sample represents a single randomly drawn plausible value for student n with ability of θ_n . Each set of 200 observations represents an individual student posterior distribution of $N(\theta_n, 0.51)$ or $N(\theta_n, 0.12)$.

Finally, 200 observations for each of the 100 students (a total of 20,000 observations) were combined to approximate the original population distributions as shown in Figure 6.9. Figure 6.9 compares the cumulative distribution functions between that recovered from the hypothesised two tests and the original generating population (black solid line). In this figure, the shorter test with a larger standard error of 0.51 is plotted as a red dashed line and the longer test with a smaller standard error of 0.12 is plotted as a green dashed line. The graph in Figure 6.9 clearly shows that with the same sample, the longer test recovered the original population very well, whereas the distribution from the shorter test is still close to a normal distribution.

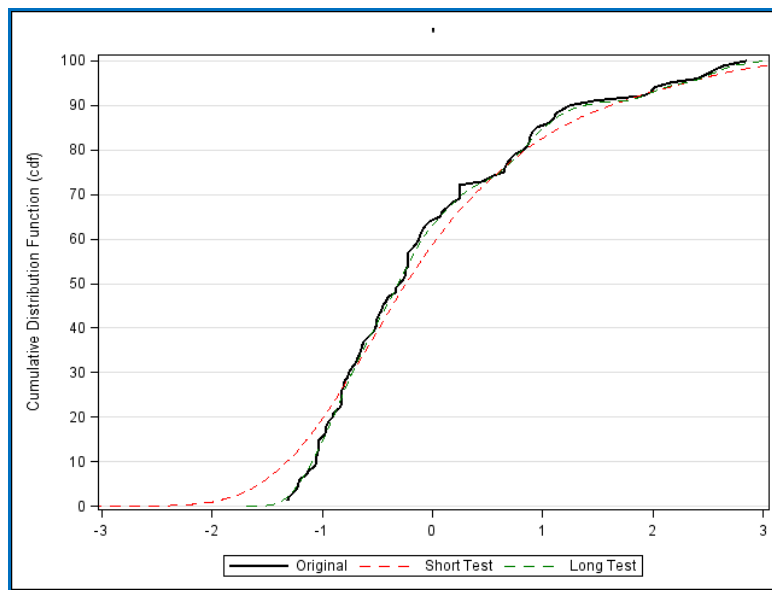
Figure 6.9 Recovered Chi-squared ($df = 5$) Distributions

Table 6.3 Simple Statistics from the Generating Distribution and the Recovered Distributions

	Standardised Chi-squared ($df = 5$)	Longer Test	Shorter Test
Mean	-0.0027	-0.0772	-0.0793
Standard Deviation	1.0208	0.9818	1.1037
Skewness	1.2903	0.9011	0.6469
Kurtosis	2.6233	0.7220	0.4757

Table 6.3 provides simple statistics for the standardised chi-squared ($df = 5$) distribution and two recovered chi-squared ($df = 5$) distributions. It shows that the recovered distribution from a longer test was more skewed and had a higher kurtosis than those from a shorter test. The mean and standard deviation from both recovered distributions were close to the mean and standard deviation of the original distribution. Figure 6.9 reveals that with all other conditions being equal, student estimates from a longer test, which had a smaller standard error of ability estimates, recovered the original population distribution better than a shorter test which had a larger standard error of ability estimates. In other

words, when a test is long enough, the prior distribution has very little impact on the posterior distribution.

6.2.5 The Effect of Sample Size

The fifth research question of interest in this dissertation is:

How does sample size affect the population parameter estimates using plausible values?

Simulation one showed that sample size had a relative large effect on the parameter estimates with the simulated conditions. That is, the larger the sample, the smaller the bias, the RMSEs and the standard errors of parameter estimates became. This pattern was also observed from the two NAPLAN empirical data studies, jurisdictions with a larger number of students had smaller standard errors of parameter estimates. The standard error refers to the variability around sample statistics, such as mean and variation. It is well known that the standard error is inversely proportional to the sample size (Kirk 1999). This implies that in practice, the aim should be to have as large a sample as practically possible.

6.3 Should One Be Concerned with the Bias and RMSE?

Both simulation studies showed there were small biases in the parameter estimates. When the sample size was small ($N = 200$), the magnitudes of the bias tended to be larger. This leads to the question: should we be worried about these biases in practice? First, it should be noted that the magnitude of the bias in all the parameter estimates was small (refer to Appendix 3). For a sample drawn from a normally distributed population, with on-target tests of different lengths (between 10 to 80 items), the bias in the estimated mean from a small sample of 200 students ranged between 0.0039 to 0.0064; the bias reduced to 0.0017 or less when the sample size increased to 2,000; and further reduced to 0.0002 or less when the sample size increased to 10,000. However, there were some large biases

observed, most often occurred in high percentile estimates. Second, in order to facilitate the interpretation of results, for most of the large-scale assessments (if not all), the reporting of results are based on scale scores not the logits (Adams & Wu, 2002; ACARA, 2015b). A scale score is normally a linear transformation from a logit value. Often this linear transformation involves a multiplying factor to have scale scores in larger number ranges. For example, students' ability estimates in logits are transformed to the NAPLAN scale scores according to the Equation 6.8.

$$\text{NAPLAN Scale Score} = A \times \text{Logit} + B \quad \text{Equation 6.8}$$

where for the NAPLAN Numeracy, $A = 60.2$ and $B = 451.3$ (ACARA, 2015b)

A bias of 0.002 logits (which is close to the bias in the estimated mean from a sample with 2,000 cases from the simulation studies) equals 0.1 of a scale score which is the magnitude of the rounding error (NAPLAN reports results with one decimal).

The RMSEs of the parameter estimates from the simulations showed there was a decreasing trend in the RMSEs with an increase in the number of plausible values used. Section 6.2.2 demonstrated that the reduction in the estimated standard errors with an increase in the number of plausible values used is related to the adjustment factor ($1/M$) in the estimated standard errors formula (Equation 2.28). Also the simulation results showed that the reduction in the estimated standard error became smaller as the number of plausible values increased. However, Table 6.2 shows that with an increase in the number of plausible values used, the standard errors of the mean estimates were actually further away from the observed standard errors. Therefore, it can be argued that the reduction in the RMSEs with an increase in the number of plausible values used may not be desirable.

Chapter 7 Conclusions and Recommendations

The use of plausible values to estimate and report population characteristics has been widely used in most large-scale education surveys or assessments since the 1980s. When the plausible value method was initially implemented, five plausible values were considered sufficient to estimate the underlying population parameters. However, in recent years, there have been some surveys, such as PISA 2015 and PIAAC 2012, which have increased the number of plausible values available for secondary analyses from 5 to 10, and in the case of NEAP 2016 from 5 to 20.

The studies in this dissertation investigated the impact of using different numbers of plausible values on the population parameter estimates under various conditions. The conclusions listed below are based on two simulation studies and two empirical data studies. The design of the simulation studies investigated the effects by varying four factors. It is not feasible and nor was it the intention of this study to cover all conditions and complexities that could be faced in real-world applications. Caution is recommended when extending the conclusions from this study into practice.

7.1 Conclusions

The most important research question to be answered with this study was: *How many plausible values are needed to produce plausible results in terms of population parameter estimates, standard error estimates and for secondary analysis?* From the simulation studies, it is shown that increasing the number of plausible values from 1 to 20 did not show evidence of improvement in the precision of parameter estimates, as there were no reduction in the bias in the parameter estimates observed for the simulated conditions. When the number of plausible values was increased, a trend for small decreases in the RMSEs and in the standard errors was observed for all parameter estimates.

The total standard error combines sampling error and measurement error in the context of this dissertation. According to Rubin's (1987) formula, the measurement error is estimated by applying an adjustment factor for a small number of plausible values ($1+1/M$) to the variance among plausible values, where M is the number of plausible values. From the simulation studies, it was found that the reduction in the standard error of estimates was related to the adjustment factor. With increasing number of plausible values used, the estimated standard errors tend to move further from the expected standard error values (refer to Table 6.2 and Figure 6.4). Although the standard errors of the parameter estimates as calculated using Rubin's formula reduced with an increase in the number of plausible values used, this may not reflect true situation in practice.

The studies with the NAPLAN 2015 Numeracy data showed that there was no consistent pattern observed in the parameter estimates or their standard errors, when using from 1 to 46 plausible values. This implies that increasing the number of plausible values, up to 46, did not improve the estimates of NAPLAN parameters, nor the standard errors of the parameter estimates. However, when sample sizes were small ($N < 2,000$), the parameter estimates using one or three plausible values showed a large difference from the estimates using five or more plausible values.

The other focus of this research was to investigate how the test reliability, in terms of test length and test targeting, population distribution and sample size related to the population parameter estimates. The two simulation studies found that the test length had a moderate impact on the population parameter estimates for the simulated conditions. Increasing test length, while other conditions are kept the same, resulted in a smaller bias and RMSEs of the population estimates regardless of the population distribution. This means that the estimation of population parameters is improved with an increase in test length.

The simulation studies also found that test targeting conditions could influence the estimates of the population parameters. The population parameter estimates using three plausible values were the similar to the estimates using five or more plausible values when a test was moderately off-target by up to 1.0 logit. Within each of the test targeting

conditions increasing the number of plausible values did not improve the population parameter estimates. When a test was severely off-target by 2.5 logits, however, the estimations had a large bias, particularly the mean and variance estimates for a skewed population distribution. This finding implies that a test should be developed to align with the test takers' ability range. When a population contains sub-populations with substantial proficiency differences, the estimates of sub-population parameters could be compromised. For example, a reasonably targeted NAPLAN test can be seriously off-target for the Northern Territory's Indigenous students which could lead to biased estimations. According to the NAPLAN 2015 National report (ACARA, 2015a), the achievement of Indigenous students in the Northern Territory was, on average, 108 scale scores (or about 1.8 logits) lower than the achievement of non-Indigenous students in Year 3 Numeracy. This means a well-targeted Numeracy test for average non-Indigenous Year 3 students would be too difficult for the Indigenous students in the Northern Territory, and may result biased estimation for the Northern Territory Indigenous students.

The studies also showed that among the four population distributions used in Simulation One, the sample drawn from a normally distributed population provided the smallest bias in the parameter estimates. The standard errors of the 95th and 97.5th percentile estimates from a chi-squared ($df = 5$) distribution were higher when compared to the standard errors from other distributions. Other parameter estimates were approximately the same among the four different distributions across various simulated conditions. There was no relation found between the number of plausible values used and the parameter estimates regardless of the population distributions.

The results from the simulation studies and the empirical data studies show that the sample size has a larger impact on the estimations of population parameters than an increase in the number of plausible values used. As expected, increasing the sample size will also improve the population estimates due to the reduction in sampling error.

The results of these investigations showed that the current practice of using five plausible values for estimating population parameters in large-scale assessments is adequate. The

reduction in the standard errors of parameter estimates with an increase in the number of plausible values was due to the application of an adjustment factor and may not be desirable. Increasing the number of plausible values to more than five showed little improvement in the population parameter estimates for the conditions used in these studies. The studies also showed that sample size has a large effect on the estimations of population parameters and increasing test length provided some benefit in improving the population parameter estimates.

7.2 Recommendation for Future Work

Based on the finding of this research, it is recommended that the current practice of using five plausible values to report NAPLAN results is appropriate and should continue. There is no evidence that increasing the number of plausible values to more than five improves population parameter estimates significantly. However, increased sample size and improved test reliability, through increased test length and better test targeting conditions, will improve the accuracy of the population parameter estimates. When sample size was small ($N < 1,000$), the analysis of the NAPLAN data showed that the variation in the parameter estimates becomes large, even when 46 plausible values were used. This means that any conclusions based on the parameter estimates from a small sample should be interpreted with extra caution.

Based on the results of this dissertation, the practice in large-scale surveys of using five plausible values for population parameter estimates and for secondary analysis is sound under certain conditions; when the assessment has a good reliability, that is, a test has 20 or more items, and when the difference between the average item difficulties and the average student abilities is within a ± 1 logit range and the sample size is reasonably large ($N > 2,000$). Using more than five plausible values provides little benefit to the estimates of population parameter estimates. When a test is extremely off-target, or the test is very short, or the sample size is very small, increasing the number of plausible values used to estimate the population parameters does not appear to compensate for the loss of precision

and accuracy of the estimates. To further confirm this observation the following additional studies are recommended:

For this dissertation, the integration procedure for MML estimation is the Monte-Carlo integration method. It is worthy of further research to investigate other integration methods, especially the Markov Chain Monte Carlo (MCMC) (van Ravenzwaaij, Cassey & Brown, 2018) method.

As this study was limited to dichotomously scored items and response data were fitted to the simple logistic model, it is recommended that studies be carried out with different item response models, including models for polytomous items, and in multidimensional contexts.

The simulation studies were carried out with complete response data where there were no missing responses in the data. Studies that include missing item response data to simulate the booklet rotation design that is widely used in many large-scale surveys (Martin *et al*, 1996; Mislevy *et al*, 1992b), would be useful.

The simulations covered test lengths from 10 items up to 80 items. In most assessments, student achievements are not only reported for main scales, but also reported for subscales which may consist of items as few as four or five items. Studies that simulate tests with very few items would be valuable.

The three covariates used in Simulation Two were categorical variables. Consideration should be given to continuous variables as covariates in the data generation. In addition, it would be helpful to investigate whether having larger proportion of variance explained by the covariates in the regression model would influence the number of plausible values needed.

Finally, NAPLAN 2018 introduced a three-stage adaptive online test. With this design, depending on the performance on the first stage testlet (ACER, 2013), a student is

assigned a more targeted testlet according to the predefined branching rule. Similarly, the third stage testlet assignment is based on the combined student performance on the first and second stages. This multi-stage test design will result in great improvement in test targeting conditions (ACER, 2013). As the nature of multi-stage adaptive test, the abilities of students within each test path will be more uniformly distributed compared to the ability distribution from the traditional paper test. It will be interesting to investigate if there is any improvement on the population parameter estimates and how the design will affect the requirement in the number of plausible values used.

References

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. United States Department of Commerce, National Bureau of Standards (NBS).

ACER (2010). *NAPLAN 2009 Technical Report*. Unpublished Report. Melbourne, Australia.

ACER (2013). *Analytical Report: Psychometric Analysis for the Trial of the Tailored Test Design*. Australian Council for Educational Research, Melbourne, December. Prepared for the Australian Curriculum, Assessment and Reporting Authority.

ACARA. (2015a). Australian Curriculum, Assessment and Reporting Authority 2015, *NAPLAN Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2015*, ACARA, Sydney

ACARA. (2015b). Australian Curriculum, Assessment and Reporting Authority 2015 *National Assessment Program – Literacy and Numeracy 2015: Technical Report*, ACARA, Sydney

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31, 162-172

Adams, R. J., Wilson, M. & Wu, M. (1997a). Multilevel Item Response Models: An Approach to Errors in Variables Regression. *Journal of Educational and Behavioural Statistics* 22, 47-76. [https://doi: 10.3102/10769986022001047](https://doi.org/10.3102/10769986022001047).

Adams, R. J., Wu, M. L. & Macaskill, G. (1997b). Scaling methodology and procedures for the mathematics and science scales. *TIMSS Technical Report Volume II: Implementation and Analysis*. Boston College, Chestnut Hill, MA.

Adams, R. J., Wilson, M., & Wang, W. C. (1997c). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1). doi:<https://doi.org/10.1177/0146621697211001>

Adams, R. J., & Wu, M. (2002). *PISA 2000 technical report*. Organisation for Economic Cooperation and Development. Pairs.

Adams, R., & Wu, M. (2007). The MRCML - a generalized Rasch model *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 58-75). New York: Springer.

Adams, R. J., Wu, M. L. & Wilson, M. R. (2015). *ACER ConQuest: Generalised item response modelling software* [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.

References

- Ahn, S., & Fessler, J. A. (Producer). (2003). Standard errors of mean, variance, and standard. <http://www.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf>.
- Allen, N. J., Johnson, E. G., Mislevy, R. J., & Thomas, N. (1999). Chapter 11 Scaling Procedure *The NAEP 1996 Technical Report*: Education Publications Center (ED Pubs).
- Allison, P. D. (2002). *Quantitative Applications in the Social Sciences: Missing data*: Thousand Oaks, CA: SAGE Publications, Inc. . doi:10.4135/9781412985079
- Andersen, E. B. (1973). Conditional inference in multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Axelrod, R. (2003). Advancing the Art of Simulation in the Social Science. *Journal of the Japanese Society for Management Information Systems*, 12, 3-16.
- Baker, F. B. (1992). *Item response theory: parameter estimation technique*. New York: Marcel Dekker, Inc.
- Beaton, A. E. (1987). *Implementing the new design: The NAEP 1983-84 technical report*. Washington, DC, Centre for Statistics (OERI/ED).
- Beaton, A. E., & Barone, J. L. (2017). Large-Scale Group-Score Assessment. In R. E. Bennett & M. v. Davier (Eds.), *Advanced Human Assessment: The Methodological, Psychological, and Policy Contributions of ETS*: Springer.
- Berezner, A., & Adams, R. J. (2017). Why large-scale assessments use scaling and item response theory. In P. Lietz, J. C. Cresswell, K. F. Rust, & R. J. Adams (Eds.), *Implementation of Large-Scale Education Assessments*: John Wiley & Sons Ltd.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, Winter 1997, 21-33.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35(2), 179-197.
- Bodner, T. E. (2008). What Improves with Increased Missing Data Imputations? *Structural Equation Modeling*, 15, 651-675. doi:0.1080/10705510802339072

References

- Box, G. E. P., & Muller, M. E. (1958). A Note on the Generation of Random Normal Deviates. *Annals Math. Statist*, 29 No. 2, 610-611. doi:10.1214/aoms/1177706645
- Brown, M. B., & Wolfe, R. A. (1983). Estimation of the variance of percentile estimates. *Computational Statistics & Data Analysis*, 167-174.
- Castro, M. G., J. (2000). Application of hierarchical linear models to meta-analysis: Study of the Monte Carlo simulation on the functioning of traditional and empirical-Bayes effect size. *Quality and Quantity*, 34(1), 33-50.
- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 330-351.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1: 1-38.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.
- Efron, B., & Stein, C. (1981). The Jackknife estimate of variance. *The Annals of Statistics*, 9(3), 586-596.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory For Psychologists*. New Jersey.
- Fisher, R. A. (1925). *Theory of statistical estimation*. Paper presented at the Proceedings of the Cambridge Philosophical Society.
- Fisher, R. A., & Yates, F. (1938). Statistical tables for biological, agricultural and medical research. New York: Hafner.
- Gardner, P. L. (1970). Test Length and the Standard Error of Measurement. *Journal of Educational Measurement*, 7(4), 271-273.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8:206–213.
- Kirk, R. E. (1999). *Statistics: An Introduction* (4th ed.): Harcourt Brace College Publishers.
- Koehlera, E., Browna, E., & Haneusea, S. (2009). On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses. *The American Statistician*, 63(2), 155-162. doi:10.1198/tast.2009.0030
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis In L. G. S. A. Stouffer, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A.

References

- Clausen. (Ed.), *Measurement and Prediction* (pp. 362-412). Princeton, NJ: Princeton University Press.
- Lietz, P., Cresswell, J. C., Rust, K. F., & Adams, R. J. (2017). *A Review of international large-scale assessment in education: Assessing component skill and collecting contextual data*. Paris: The World Bank, Washington, D. C./OECD Publishing.
- Linacre, J. M. (1993). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *American Statistician*, 37, 218-220.
- Longford, N. T. (2005). *Missing Data and Small-Area Estimation: Modern analytical equipment for the survey statistician*. New York: Springer.
- Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological* (pp. 13, 517-548).
- Lord, F. M. (1959). Tests of the same length do have the same standard error of measurement. *Educational and Psychological Measurement*, 19, 233-239.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Pub. Co.
- Luo, Y., & Dimiter M, D. (2019). A Short Note on Obtaining Point Estimates of the IRT Ability Parameter With MCMC Estimation in MPlus: How Many Plausible Values Are Needed? *Educational and Psychological Measurement*, 79(2), 272-287. doi:10.1177/0013164418777569.
- Marsman, M., Maris, G., Bechger, T., & Glas, C. (2016). What can we learn from plausible values? *Psychometrika*, 81, no. 2, 274–289.
- Martin, M. O., & Kelly, D. L. E. (1996). *Third international mathematics and science study technical report volume I: design and development*. Chestnut Hill, MA: Boston College.
- Martin, M. O., Mullis, I. V. S. and Kennedy, A. M. (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika* 49(3): 359-381.
- Mislevy, R. J. (1985). *Inferences about latent populations from complex samples*. Princeton, New Jersey.

References

- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika* 56(2): 177-196.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992a). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement* 29: 133-161.
- Mislevy, R. J., & Bock, R. D. (1983). BILOG: Item analysis and test scoring with binary logistic models [Computer program]. Mooresville IN: Scientific Software. Inc.
- Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992b). Scaling procedures in NAEP. *Journal of Educational Statistics*, 17, 131-154. <https://doi:10.3102/10769986017002131>.
- Monseur, C., & Adams, R. (2009). Plausible values: How to deal with their limitations. *Journal of Applied Measurement*, 10(3), 320-334.
- Murray, T. S., Kirsch, I. S. & Jenkins, L. (1998). *Adult literacy in OECD countries: Technical report on the first international adult literacy survey*. Washington, DC, U.S. Department of Education, National Center for Education Statistics.
- National Assessment of Educational Progress. *Creation of Plausible Values* Retrieved October 20, 2016, from https://nces.ed.gov/nationsreportcard/tdw/analysis/est_pv_creation.aspx.
- Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1-32.
- OECD (2013). *Technical report of the survey of adult skills (PIAAC)*, Retrieved from [http://www.oecd.org/site/piaac/ Technical%20Report_17OCT13.pdf](http://www.oecd.org/site/piaac/Technical%20Report_17OCT13.pdf).
- OECD (Producer). (2017, July). PISA 2015 technical report. www.oecd.org/pisa/data/2015-technical-report.
- OECD. PISA 2018 technical report <https://www.oecd.org/pisa/pisa-for-development/pisafordevelopment2018technicalreport/>
- Pan, Q., & Wei, R. (2016). Fraction of Missing Information (γ) at Different Missing Data Fractions in the 2012 NAMCS Physician Workflow Mail Survey. *Applied Mathematics*, 7, 1057-1067. doi:10.4236/am.2016.710093
- Peng, R. (2018). Advanced Statistical Computing, <https://bookdown.org/rdpeng/advstatcomp/>
- Pinheiro, J. C. & Chao, E. C. (2006) Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models, *Journal of Computational and Graphical Statistics*, 15:1, 58-81, DOI: 10.1198/106186006X96962

References

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). "Section 4.7. Adaptive Quadrature" *Numerical Recipes: The Art of Scientific Computing (3rd ed.)*, New York: Cambridge University Press. ISBN 978-0-521-88068-8.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published 1960).

Reckase, M. D. (2009). Multidimensional Item Response Theory Models. In: Multidimensional Item Response Theory. *Statistics for Social and Behavioral Sciences*: Springer, New York, NY.

Rogers, A., Tang, C., Lin, M.-J., & Kandathil, M. (2006). DGroup (Computer software). Princeton, NJ: Educational Testing Service.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, Wiley.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Schafer, J. L. and M. K. Olsen (1998). Multiple Imputation for Multivariate Missing-Data Problem: A Data Analyst's Perspective. *Multivariate Behavioural Research* 33:4, 545-571. https://doi:10.1207/s15327906mbr3304_5.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen., J. A. (1950). *Measurement and Prediction*. Princeton, NJ: Princeton University Press.

SAS®/STA software. SAS Institute Inc., Cary, NC: Release 9.3, 2015.

Sokolowski, J. A., & Banks, C. M. (2009). *Principles of Modeling and Simulation*: Hoboken, NJ: Wiley.

Thissen, D. (1982). Marginal Maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175-186.

Thomson, S., De Bortoli, L., & Underwood, C. (2017). *PISA 2015: Reporting Australia's results*. 19 Prospect Hill Road, Camberwell, Victoria, 3124, Australia: Australian Council for Educational Research Ltd.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, 11, 1-13.

van Belle, G. (2008). *Statistical rules of thumb*. New Jersey: Hoboken, N.J.: Wiley, c2008.

van der Linden, W. J. (2016). *Handbook of Item Response Theory* (Vol. one Models). 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742: CRC Press, Taylor & Francis Group, LLC.

References

- van Ravenzwaaij, D., Cassey, P. & Brown, S.D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychon Bull Rev* 25, 143–154 (2018) doi:10.3758/s13423-016-1015-8
- Volodin, N. and Adams, R. J. (1995). *Identifying and Estimating a D-dimensional Rasch model*. Camberwell, Victoria, Australia, Australian Council for Educational Research.
- von Davier, M. (2016). Private communication.
- von Davier, M., Gonzalez, E. J., & Mislevy, R. J. (2009). What are Plausible Values and Why are They Useful? *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 2, 9-36.
- von Hippel, P. T. (2018). How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule *Sociological Methods and Research*. doi:<https://doi.org/10.1177/0049124117747303>.
- Warm, T. A. (1985). Weighted maximum likelihood estimation of ability in item response theory with tests of finite length. Technical Report CGI-TR-85-08. Oklahoma City, U.S. Cost Guard Institute.
- Wilson, M. (2005). *Constructing Measures. An Item Response Modeling Approach*: Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23-48.
- Wu, M. L. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31, 114-128.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest: Generalised item response modeling software* [Computer software]. Version 2.0 Camberwell, Victoria: Australian Council for Educational Research.
- Yamamoto, K. K., E. (2000). Scaling Methodology and Procedures for the TIMSS Mathematics and Science Scales. *TIMSS 1999 Technical Report*. MA: (J. and K. 1997) Boston College.
- Zhang, J., Isham, S. P., & Worthington, L. H. (2001). Data analysis of the national reading assessment. In N. L. Allen, J. R. Donoghue, & T. L. Schoeps (Eds.), *The NAEP 1998 Technical Report* (pp. 283-305). Washington, DC: National Center for Education Statistics.

Appendix 1 Item Parameters for the Simulation Studies

10 items

Item Number	Test Targeting Condition			
	On Target	Off-target (0.5)	Off-target (1.0)	Off-target (2.5)
1	2.33333	2.83333	3.33333	4.83333
2	-0.33333	0.16667	0.66667	2.16667
3	-3.00000	-2.50000	-2.00000	-0.50000
4	-1.00000	-0.50000	0.00000	1.50000
5	0.33333	0.83333	1.33333	2.83333
6	1.00000	1.50000	2.00000	3.50000
7	-1.66667	-1.16667	-0.66667	0.83333
8	3.00000	3.50000	4.00000	5.50000
9	1.66667	2.16667	2.66667	4.16667
10	-2.33333	-1.83333	-1.33333	0.16667

20 items

Item Number	Test Targeting Condition			
	On Target	Off-target (0.5)	Off-target (1.0)	Off-target (2.5)
1	1.73684	2.23684	2.73684	4.23684
2	0.15789	0.65789	1.15789	2.65789
3	-3.00000	-2.50000	-2.00000	-0.50000
4	-0.15789	0.34211	0.84211	2.34211
5	0.47368	0.97368	1.47368	2.97368
6	1.10526	1.60526	2.10526	3.60526
7	-0.47368	0.02632	0.52632	2.02632
8	2.68421	3.18421	3.68421	5.18421
9	1.42105	1.92105	2.42105	3.92105
10	-2.36842	-1.86842	-1.36842	0.13158
11	0.78947	1.28947	1.78947	3.28947
12	-2.05263	-1.55263	-1.05263	0.44737
13	2.36842	2.86842	3.36842	4.86842
14	-2.68421	-2.18421	-1.68421	-0.18421
15	-1.73684	-1.23684	-0.73684	0.76316
16	3.00000	3.50000	4.00000	5.50000
17	2.05263	2.55263	3.05263	4.55263
18	-0.78947	-0.28947	0.21053	1.71053
19	-1.42105	-0.92105	-0.42105	1.07895
20	-1.10526	-0.60526	-0.10526	1.39474

40 items

Item Number	Test Targeting Condition			
	On Target	Off-target (0.5)	Off-target (1.0)	Off-target (2.5)
1	1.15385	1.65385	2.15385	3.65385
2	-0.53846	-0.03846	0.46154	1.96154
3	-3.00000	-2.50000	-2.00000	-0.50000
4	-1.15385	-0.65385	-0.15385	1.34615
5	0.23077	0.73077	1.23077	2.73077
6	0.53846	1.03846	1.53846	3.03846
7	-1.30769	-0.80769	-0.30769	1.19231
8	2.84615	3.34615	3.84615	5.34615
9	1.00000	1.50000	2.00000	3.50000
10	-2.53846	-2.03846	-1.53846	-0.03846
11	0.38462	0.88462	1.38462	2.88462
12	-2.38462	-1.88462	-1.38462	0.11538
13	2.69231	3.19231	3.69231	5.19231
14	-2.69231	-2.19231	-1.69231	-0.19231
15	-2.23077	-1.73077	-1.23077	0.26923
16	3.00000	3.50000	4.00000	5.50000
17	1.30769	1.80769	2.30769	3.80769
18	-1.46154	-0.96154	-0.46154	1.03846
19	-1.92308	-1.42308	-0.92308	0.57692
20	-1.61538	-1.11538	-0.61538	0.88462
21	0.07692	0.57692	1.07692	2.57692
22	2.53846	3.03846	3.53846	5.03846
23	-2.84615	-2.34615	-1.84615	-0.34615
24	-0.23077	0.26923	0.76923	2.26923
25	1.92308	2.42308	2.92308	4.42308
26	1.61538	2.11538	2.61538	4.11538
27	1.76923	2.26923	2.76923	4.26923
28	-1.76923	-1.26923	-0.76923	0.73077
29	0.84615	1.34615	1.84615	3.34615
30	1.46154	1.96154	2.46154	3.96154
31	-0.07692	0.42308	0.92308	2.42308
32	0.69231	1.19231	1.69231	3.19231
33	2.23077	2.73077	3.23077	4.73077
34	-0.38462	0.11538	0.61538	2.11538
35	-2.07692	-1.57692	-1.07692	0.42308
36	-0.84615	-0.34615	0.15385	1.65385
37	-0.69231	-0.19231	0.30769	1.80769
38	2.07692	2.57692	3.07692	4.57692
39	-1.00000	-0.50000	0.00000	1.50000
40	2.38462	2.88462	3.38462	4.88462

60 Items

Item Number	Test Targeting Condition			
	On Target	Off-target (0.5)	Off-target (1.0)	Off-target (2.5)
1	1.37288	1.87288	2.37288	3.87288
2	-0.45763	0.04237	0.54237	2.04237
3	-2.89831	-2.39831	-1.89831	-0.39831
4	-1.06780	-0.56780	-0.06780	1.43220
5	0.15254	0.65254	1.15254	2.65254
6	0.55932	1.05932	1.55932	3.05932
7	-1.16949	-0.66949	-0.16949	1.33051
8	2.89831	3.39831	3.89831	5.39831
9	1.16949	1.66949	2.16949	3.66949
10	-2.49153	-1.99153	-1.49153	0.00847
11	0.35593	0.85593	1.35593	2.85593
12	-2.38983	-1.88983	-1.38983	0.11017
13	2.79661	3.29661	3.79661	5.29661
14	-2.59322	-2.09322	-1.59322	-0.09322
15	-2.18644	-1.68644	-1.18644	0.31356
16	3.00000	3.50000	4.00000	5.50000
17	1.47458	1.97458	2.47458	3.97458
18	-1.27119	-0.77119	-0.27119	1.22881
19	-1.88136	-1.38136	-0.88136	0.61864
20	-1.57627	-1.07627	-0.57627	0.92373
21	-0.05085	0.44915	0.94915	2.44915
22	2.69492	3.19492	3.69492	5.19492
23	-2.79661	-2.29661	-1.79661	-0.29661
24	-0.25424	0.24576	0.74576	2.24576
25	2.18644	2.68644	3.18644	4.68644
26	1.77966	2.27966	2.77966	4.27966
27	2.08475	2.58475	3.08475	4.58475
28	-1.77966	-1.27966	-0.77966	0.72034
29	1.06780	1.56780	2.06780	3.56780
30	1.67797	2.17797	2.67797	4.17797
31	-0.15254	0.34746	0.84746	2.34746
32	0.66102	1.16102	1.66102	3.16102
33	2.38983	2.88983	3.38983	4.88983
34	-0.35593	0.14407	0.64407	2.14407
35	-2.08475	-1.58475	-1.08475	0.41525
36	-0.66102	-0.16102	0.33898	1.83898
37	-0.55932	-0.05932	0.44068	1.94068
38	2.28814	2.78814	3.28814	4.78814
39	-0.96610	-0.46610	0.03390	1.53390
40	2.49153	2.99153	3.49153	4.99153
41	0.45763	0.95763	1.45763	2.95763
42	0.86441	1.36441	1.86441	3.36441
43	-2.69492	-2.19492	-1.69492	-0.19492
44	-1.67797	-1.17797	-0.67797	0.82203
45	-0.86441	-0.36441	0.13559	1.63559
46	-3.00000	-2.50000	-2.00000	-0.50000
47	-1.37288	-0.87288	-0.37288	1.12712
48	1.88136	2.38136	2.88136	4.38136
49	1.57627	2.07627	2.57627	4.07627
50	0.25424	0.75424	1.25424	2.75424
51	-1.98305	-1.48305	-0.98305	0.51695
52	1.98305	2.48305	2.98305	4.48305
53	-1.47458	-0.97458	-0.47458	1.02542
54	0.05085	0.55085	1.05085	2.55085
55	1.27119	1.77119	2.27119	3.77119
56	0.76271	1.26271	1.76271	3.26271

Appendix 1

Item Parameters for the Simulation Studies

57	-2.28814	-1.78814	-1.28814	0.21186
58	0.96610	1.46610	1.96610	3.46610
59	-0.76271	-0.26271	0.23729	1.73729
60	2.59322	3.09322	3.59322	5.09322

80 Items

Item Number	Test Targeting Condition			
	On Target	Off-target (0.5)	Off-target (1.0)	Off-target (2.5)
1	1.48101	1.98101	2.48101	3.98101
2	-0.18987	0.31013	0.81013	2.31013
3	-2.92405	-2.42405	-1.92405	-0.42405
4	-1.10127	-0.60127	-0.10127	1.39873
5	0.34177	0.84177	1.34177	2.84177
6	0.72152	1.22152	1.72152	3.22152
7	-1.17722	-0.67722	-0.17722	1.32278
8	2.84810	3.34810	3.84810	5.34810
9	1.32911	1.82911	2.32911	3.82911
10	-2.54430	-2.04430	-1.54430	-0.04430
11	0.56962	1.06962	1.56962	3.06962
12	-2.46835	-1.96835	-1.46835	0.03165
13	2.69620	3.19620	3.69620	5.19620
14	-2.62025	-2.12025	-1.62025	-0.12025
15	-2.31646	-1.81646	-1.31646	0.18354
16	2.92405	3.42405	3.92405	5.42405
17	1.55696	2.05696	2.55696	4.05696
18	-1.32911	-0.82911	-0.32911	1.17089
19	-2.08861	-1.58861	-1.08861	0.41139
20	-1.70886	-1.20886	-0.70886	0.79114
21	0.18987	0.68987	1.18987	2.68987
22	2.62025	3.12025	3.62025	5.12025
23	-2.84810	-2.34810	-1.84810	-0.34810
24	-0.03797	0.46203	0.96203	2.46203
25	2.16456	2.66456	3.16456	4.66456
26	1.78481	2.28481	2.78481	4.28481
27	2.08861	2.58861	3.08861	4.58861
28	-1.93671	-1.43671	-0.93671	0.56329
29	1.25316	1.75316	2.25316	3.75316
30	1.70886	2.20886	2.70886	4.20886
31	0.11392	0.61392	1.11392	2.61392
32	0.79747	1.29747	1.79747	3.29747
33	2.31646	2.81646	3.31646	4.81646
34	-0.11392	0.38608	0.88608	2.38608
35	-2.24051	-1.74051	-1.24051	0.25949
36	-0.41772	0.08228	0.58228	2.08228
37	-0.34177	0.15823	0.65823	2.15823
38	2.24051	2.74051	3.24051	4.74051
39	-1.02532	-0.52532	-0.02532	1.47468
40	2.46835	2.96835	3.46835	4.96835
41	0.64557	1.14557	1.64557	3.14557
42	1.02532	1.52532	2.02532	3.52532
43	-2.77215	-2.27215	-1.77215	-0.27215
44	-1.78481	-1.28481	-0.78481	0.71519
45	-0.87342	-0.37342	0.12658	1.62658
46	-3.00000	-2.50000	-2.00000	-0.50000
47	-1.40506	-0.90506	-0.40506	1.09494
48	1.86076	2.36076	2.86076	4.36076
49	1.63291	2.13291	2.63291	4.13291
50	0.49367	0.99367	1.49367	2.99367
51	-2.16456	-1.66456	-1.16456	0.33544
52	1.93671	2.43671	2.93671	4.43671
53	-1.63291	-1.13291	-0.63291	0.86709
54	0.26582	0.76582	1.26582	2.76582
55	1.40506	1.90506	2.40506	3.90506
56	0.87342	1.37342	1.87342	3.37342

57	-2.39241	-1.89241	-1.39241	0.10759
58	1.17722	1.67722	2.17722	3.67722
59	-0.79747	-0.29747	0.20253	1.70253
60	2.54430	3.04430	3.54430	5.04430
61	-0.49367	0.00633	0.50633	2.00633
62	-0.72152	-0.22152	0.27848	1.77848
63	2.01266	2.51266	3.01266	4.51266
64	2.39241	2.89241	3.39241	4.89241
65	3.00000	3.50000	4.00000	5.50000
66	1.10127	1.60127	2.10127	3.60127
67	-2.69620	-2.19620	-1.69620	-0.19620
68	-1.48101	-0.98101	-0.48101	1.01899
69	0.41772	0.91772	1.41772	2.91772
70	-0.94937	-0.44937	0.05063	1.55063
71	0.03797	0.53797	1.03797	2.53797
72	-1.55696	-1.05696	-0.55696	0.94304
73	-1.86076	-1.36076	-0.86076	0.63924
74	-0.26582	0.23418	0.73418	2.23418
75	2.77215	3.27215	3.77215	5.27215
76	-0.56962	-0.06962	0.43038	1.93038
77	-2.01266	-1.51266	-1.01266	0.48734
78	-1.25316	-0.75316	-0.25316	1.24684
79	0.94937	1.44937	1.94937	3.44937
80	-0.64557	-0.14557	0.35443	1.85443

Appendix 2 ConQuest Commands for the Simulation Studies and the Empirical Studies

Simulation One

```
let npersons=2; /* No. of cases 100: 2, 20 & 100 */
let nitems=60; /* No. of items: 10,20, 40, 60 & 80 */
let path=C:\Sim; /* File path */
let case=NorMix; /* Sample Distributions used in file name: Normal, Chi5 & Uniform */
let ability=NormalMix(-0.894:0.2:0.894:0.2:0.5); /*Sample distributions: Normal(0:1); ChiSq(5) & uniform(-
1.732:1.732) */
let target=F3; /* Targeting: On, F1(=0.5), F2(=1.0) & F3(=2.5) */

/***** Following section is to generate data files *****/
dofor i=1-1000; /* Replication Number */
export logfile >> %path%\results\%nitems%_%npersons%_%i%.log;
set seed=date, warnings=no;
generate !nitems=%nitems%, npersons=%npersons%00, maxscat=1,centre=no,
itemdist=%path%\ItemParameterFiles\%target%_%nitems%.itm, abilitydist=%ability%
>> %path%\Results\data\dummy%target%_%nitems%_%i%.dat;

/***** calibrate WITHOUT regression *****/
set warnings=no,constraint=cases,n_plausible=54;
datafile %path%\Results\data\dummy%target%_%nitems%_%i%.dat;
format responses 1-%nitems% ;

model item ;

estimate! converge=0.0001,abilities=yes,fit=no,stderr=quick;
export parameters >> %path%\results\%target%_%nitems%_%i%.itm;
show cases ! estimates=latent,filetype=spss >> %path%\Results\PV\%target%_%nitems%_%i%.sav;
show ! estimates=latent >> %path%\Results\%target%_%nitems%_%i%.shw;

reset;
enddo;

quit;
```

Simulation Two

```

let npersons= 20; /* no. of case x100 */
let nitems=80; /* No. of items: 10,20, 40, 60 & 80 */
let path=C:\Sim; /* File path */
let case=Normal; /* Sample Distributions used in file name */
let ability=Normal(0:1); /*Sample distributions */
let target=F1; /* Targeting: On, F1(=0.5), F2(=1.0) & F3(=2.5) */
execute

/***** generate data first *****/
dofor i=1-1000; /* Replication Number */
set seed=date,warnings=no;

generate !nitems=%nitems%, npersons=%npersons%00, maxscat=1,centre=no,
itemdist=%path%\ItemParameterFiles\%target%_%nitems%.itm, abilitydist=%ability%,
regfile=%path%\Data\dummy_NT5.dat(gender:LBOTE:ATSI:Geol:Sector)
>> %path%\data\dummy%i%.dat,%path%\data\dummy.itm,%path%\data\dummy.abl;

/***** calibrate WITH regression *****/
set warnings=no, constrain=items,n_plausible=54;
export logfile >> %path%\log\%nitems%_%npersons%_%i%case.log;
datafile %path%\data\dummy%i%.dat;
format responses 1-%nitems% gender 81-90 LBOTE 91-100 ATSI 101-110 Geol 111-120 Sector 121-130;
regression gender LBOTE ATSI Geol Sector;

model item ;

estimate! iterations=1000,converge=0.00001,abilities=yes,fit=no,stderr=quick;
show cases ! estimates=latent,filetype=spss >> %path%\%target%_%nitems%_%i%case.sav;
export parameters >> %path%\results\%target%_%nitems%_%i%case.itm;
show ! estimates=latent >> %path%\Results\%target%_%nitems%_%i%case.shw;

reset;

enddo;
quit;

```

Empirical Studies

Year 3

```
Let path=C:\NAPLAN;
Let year=3;
```

```
dofor state=NSW VIC QLD WA SA TAS ACT NT;
```

```
Let NSWReg=Metro Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let VICReg=Metro Remo scC scI Fem ATSI ATSIM LBOTE;
Let QLDReg=Metro Remo VRemo GeolM scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let WAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
```

```
Let SAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let TASReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let ACTReg=Metro scC scI scH Fem ATSI ATSIM LBOTE;
Let NTReg=Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
```

```
title %state% NAPLAN 2015 - Year %year% - PV Based on Sep2015 Data;
set warnings=no,keepplastests=yes,constraints=none,n_plausible=100;
export logfile      >> %path%\Log\Y%year%N_%state%.log;
datafile  %path%\CE_3%state%_wle_check.dat;
format responses 99-133 Metro 275 Remo 276 VRemo 277 GeolM 278 Fem 279 SexM 280 ATSI 281 ATSIM 282
LBOTE 283 LBOTEM 284
scC 285 scI 286 scH 287 edu0 305 edu3 to edu6 306-309 occ1 310 occ3 311 occ4 312 occ8 313 occ9 314 WLEN
359-369;
```

```
codes 0,1,2,3,4,5,8,9,M;
```

```
Key 1144213423342421213311231124131211111;
```

```
if %state%==NSW;
Regression %NSWReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==VIC;
Regression %VICReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==QLD;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==WA;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==SA;
Regression %SAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==TAS;
Regression %TASReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==ACT;
Regression %ACTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==NT;
Regression %NTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
```

```
model item;
```

```
import anchor_parameters << %path%\CQ\Anchor_parameters_%year%N.txt;
export parameters      >> %path%\System\Year%year%N_%state%.par;
estimate! iter=1000,conv=0.0001,stderr=none,fit=no;
```

```
show cases!estimates=Latent >> %path%\PV\Year%year%N_%state%.pls;
show cases!estimates=latent,filetype=spss >> %path%\PV\Year%year%N_%state%.sav;
italan !estimates=latent >> %path%\Results\Year%year%N_%state%.itn;
italan !estimates=latent, format=summary >> %path%\Results\Year%year%N_%state%.its;
show !estimates=Latent >> %path%\Results\Year%year%N_%state%.shw;
```

```
reset;
enddo;
```

```
quit;
```

Year 5

```
let path = C:\NAPLAN;
Let year=5;
dofor state=NSW VIC QLD WA SA TAS ACT NT;
```

```
Let NSWReg=Metro Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let VICReg=Metro Remo scC scI Fem ATSI ATSIM LBOTE;
Let QLDReg=Metro Remo VRemo GeolM scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let WAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
```

```
Let SAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let TASReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let ACTReg=Metro scC scI scH Fem ATSI ATSIM LBOTE;
Let NTReg=Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
```

```
title %state% NAPLAN 2015 - Year %year% Reading - PV Based on Sep2015 Data;
set warnings=no,keepplastests=yes,constraints=none,n_plausible=100;
export logfile >> %path%\Log\Y%year%_state%.log;
datafile %path%\ForPV\CE_5%state%_wle_check.dat;
format responses 99-138
Metro 275 Remo 276 VRemo 277 GeolM 278 Fem 279 SexM 280 ATSI 281 ATSIM 282 LBOTE 283 LBOTEM
284
scC 285 scI 286 scH 287 edu0 305 edu3 to edu6 306-309 occ1 310 occ3 311 occ4 312 occ8 313 occ9 314 WLEN
359-369;
labels << Z:\PhD\2017\CQ\UniD_Y%year%N.lab;
```

```
codes 0,1,2,3,4,5,8,9,M;
```

```
Key 1423231224411431231113343141423111125211!1;
```

```
if %state%==NSW;
Regression %NSWReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==VIC;
Regression %VICReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==QLD;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==WA;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==SA;
Regression %SAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==TAS;
Regression %TASReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==ACT;
Regression %ACTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
```

```

if %state%==NT;
Regression %NTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;

model item;

import anchor_parameters << Z:\PhD\2017\CQ\Anchor_parameters_%year%N.txt;

export parameters >> %path%\System\Year%year%N_%state%.par;
estimate! iter=1000,conv=0.0001,stderr=none,fit=no;

show cases!estimates=Latent >> %path%\PV\Year%year%N_%state%.pls;
show cases!estimates=latent,filetype=spss >> %path%\PV\Year%year%N_%state%.sav;
%path%\Results\Year%year%N_%state%.its;
show !estimates=Latent >> %path%\Results\Year%year%N_%state%.shw;

reset;
enddo;

quit;

```

Year 7

```

let path = C:\NAPLAN;
Let year=7;

dofor state=NSW VIC QLD WA SA TAS ACT NT;

Let NSWReg=Metro Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let VICReg=Metro Remo scC scI Fem ATSI ATSIM LBOTE;
Let QLDReg=Metro Remo VRemo GeolM scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let WAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;

Let SAREg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let TASReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let ACTReg=scC scI scH Fem ATSI ATSIM LBOTE;
Let NTReg=Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;

title %state% NAPLAN 2015 - Year %year% - PV Based on Sep2015 Data;
set warnings=no,keepplastests=yes,constraints=none,n_plausible=100;
export logfile >> %path%\Log\Y%year%N_%state%.log;
datafile %path%\ForPV\CE_7%state%_wle_check.dat;
format responses 66-97,99-130 Metro 275 Remo 276 VRemo 277 GeolM 278 Fem 279 SexM 280 ATSI 281 ATSIM
282 LBOTE 283 LBOTEM 284
scC 285 scI 286 scH 287 edu0 305 edu3 to edu6 306-309 occ1 310 occ3 311 occ4 312 occ8 313 occ9 314 WLEN
359-369;

codes 0,1,2,3,4,5,8,9,M;
Key 224232351212144343121143341112113214121112431332423213151411441111;

if %state%==NSW;
Regression %NSWReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==VIC;
Regression %VICReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==QLD;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==WA;
Regression %WAREg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;

```

```

if %state%==SA;
Regression %SAReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==TAS;
Regression %TASReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==ACT;
Regression %ACTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==NT;
Regression %NTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;

```

```

model item;

```

```

import anchor_parameters << %path%\CQ\Anchor_parameters_%year%N.txt;
export parameters >> %path%\System\Year%year%N_%state%.par;
estimate! iter=1000,conv=0.0001,stderr=none,fit=no;

```

```

show cases!estimates=Latent >> %path%\PV\Year%year%N_%state%.pls;
show cases!estimates=latent,filetype=spss >> %path%\PV\Year%year%N_%state%.sav;
itanal !estimates=latent >> %path%\Results\Year%year%N_%state%.itn;
show !estimates=Latent >> %path%\Results\Year%year%N_%state%.shw;

```

```

reset;
enddo;

```

```

quit;

```

Year 9

```

let path = C:\NAPLAN;
Let year=9;

```

```

dofor state=NSW VIC QLD WA SA TAS ACT NT;

```

```

Let NSWReg=Metro Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let VICReg=Metro Remo scC scI Fem ATSI ATSIM LBOTE;
Let QLDReg=Metro Remo VRemo GeolM scC scI Fem ATSI ATSIM LBOTE LBOTEM;
Let WAReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;

```

```

Let SAReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let TASReg=Metro Remo VRemo scC scI scH Fem ATSI ATSIM LBOTE LBOTEM;
Let ACTReg=scC scI scH Fem ATSI ATSIM LBOTE;
Let NTReg=Remo VRemo scC scI Fem ATSI ATSIM LBOTE LBOTEM;

```

```

title %state% NAPLAN 2015 - Year %year% - PV Based on Sep2015 Data;
set warnings=no,keepplastests=yes,constraints=none,n_plausible=100;
export logfile >> %path%\Log\Y%year%N_%state%.log;
datafile %path%\ForPV\CE_9%state%_wle_check.dat;
format responses 66-97,99-130 Metro 275 Remo 276 VRemo 277 GeolM 278 Fem 279 SexM 280 ATSI 281 ATSIM
282 LBOTE 283 LBOTEM 284
scC 285 scI 286 scH 287 edu0 305 edu3 to edu6 306-309 occ1 310 occ3 311 occ4 312 occ8 313 occ9 314 WLEN
359-369;

```

```

codes 0,1,2,3,4,5,8,9,M;

```

```

Key 2313213221433313414211111441111122122423313424134114111144541311!1;

```

```

if %state%==NSW;
Regression %NSWReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==VIC;

```

```

Regression %VICReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==QLD;
Regression %WAreReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==WA;
Regression %WAreReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==SA;
Regression %SAreReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==TAS;
Regression %TASReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==ACT;
Regression %ACTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;
if %state%==NT;
Regression %NTReg% WLEN edu0 edu3 edu4 edu5 edu6 occ1 occ3 occ4 occ8 occ9;
endif;

model item;

import anchor_parameters << %path%\CQ\Anchor_parameters_%year%N.txt;
export parameters >> %path%\System\Year%year%N_%state%.par;
estimate! iter=1000,conv=0.0001,stderr=none,fit=no;

show cases!estimates=Latent >> %path%\PV\Year%year%N_%state%.pls;
show cases!estimates=latent,filetype=spss >> %path%\PV\Year%year%N_%state%.sav;
itanal !estimates=latent >> %path%\Results\Year%year%N_%state%.itn;
show !estimates=Latent >> %path%\Results\Year%year%N_%state%.shw;

reset;
enddo;

quit;

```

Appendix 3 Estimated Bias and RMSEs of Simulation One

Table 1 Bias and RMSE for Mean by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0062	0.0060	0.0040	0.0052	0.0051	0.0048	0.1093	0.1039	0.1020	0.1022	0.1014	0.1007
200	Normal	On-target	20	0.0058	0.0055	0.0064	0.0057	0.0061	0.0063	0.0935	0.0894	0.0876	0.0872	0.0869	0.0867
200	Normal	On-target	40	0.0035	0.0041	0.0046	0.0039	0.0039	0.0043	0.0839	0.0805	0.0800	0.0790	0.0790	0.0789
200	Normal	On-target	60	0.0062	0.0053	0.0056	0.0056	0.0055	0.0054	0.0785	0.0754	0.0752	0.0749	0.0748	0.0744
200	Normal	On-target	80	0.0041	0.0040	0.0039	0.0045	0.0043	0.0042	0.0747	0.0737	0.0732	0.0725	0.0727	0.0725
200	Normal	Off by 0.5	10	-0.0035	-0.0030	-0.0052	-0.0042	-0.0050	-0.0046	0.1127	0.1079	0.1063	0.1059	0.1053	0.1046
200	Normal	Off by 0.5	20	-0.0040	-0.0027	-0.0033	-0.0042	-0.0038	-0.0036	0.0964	0.0903	0.0890	0.0889	0.0886	0.0889
200	Normal	Off by 0.5	40	-0.0025	-0.0016	-0.0023	-0.0016	-0.0019	-0.0022	0.0844	0.0814	0.0810	0.0801	0.0798	0.0800
200	Normal	Off by 0.5	60	-0.0030	-0.0016	-0.0019	-0.0028	-0.0021	-0.0021	0.0797	0.0768	0.0756	0.0761	0.0755	0.0755
200	Normal	Off by 0.5	80	-0.0017	-0.0030	-0.0026	-0.0018	-0.0022	-0.0020	0.0753	0.0737	0.0732	0.0732	0.0728	0.0726
200	Normal	Off by 1.0	10	-0.0095	-0.0108	-0.0115	-0.0106	-0.0106	-0.0103	0.1291	0.1203	0.1209	0.1200	0.1195	0.1193
200	Normal	Off by 1.0	20	-0.0124	-0.0126	-0.0124	-0.0127	-0.0130	-0.0128	0.1036	0.0971	0.0969	0.0954	0.0954	0.0945
200	Normal	Off by 1.0	40	-0.0102	-0.0109	-0.0099	-0.0107	-0.0105	-0.0102	0.0861	0.0845	0.0831	0.0824	0.0824	0.0826
200	Normal	Off by 1.0	60	-0.0116	-0.0109	-0.0111	-0.0115	-0.0117	-0.0117	0.0829	0.0800	0.0782	0.0788	0.0787	0.0783
200	Normal	Off by 1.0	80	-0.0115	-0.0112	-0.0103	-0.0104	-0.0104	-0.0100	0.0779	0.0752	0.0753	0.0744	0.0746	0.0740
200	Uniform	On-target	10	0.0072	0.0054	0.0055	0.0054	0.0059	0.0055	0.1103	0.1041	0.1020	0.1022	0.1020	0.1016
200	Uniform	On-target	20	0.0051	0.0045	0.0063	0.0055	0.0058	0.0056	0.0916	0.0887	0.0870	0.0861	0.0863	0.0864
200	Uniform	On-target	40	0.0038	0.0034	0.0037	0.0040	0.0042	0.0037	0.0848	0.0812	0.0812	0.0805	0.0801	0.0799
200	Uniform	On-target	60	0.0057	0.0058	0.0062	0.0059	0.0061	0.0062	0.0807	0.0786	0.0782	0.0779	0.0773	0.0773
200	Uniform	On-target	80	0.0047	0.0049	0.0047	0.0049	0.0045	0.0047	0.0799	0.0776	0.0769	0.0766	0.0769	0.0769
200	Uniform	Off by 0.5	10	-0.0069	-0.0080	-0.0079	-0.0080	-0.0074	-0.0074	0.1160	0.1091	0.1097	0.1077	0.1072	0.1076
200	Uniform	Off by 0.5	20	-0.0048	-0.0046	-0.0052	-0.0046	-0.0055	-0.0052	0.0937	0.0899	0.0887	0.0875	0.0871	0.0876
200	Uniform	Off by 0.5	40	-0.0051	-0.0051	-0.0039	-0.0042	-0.0045	-0.0044	0.0861	0.0832	0.0820	0.0817	0.0816	0.0818
200	Uniform	Off by 0.5	60	-0.0040	-0.0027	-0.0028	-0.0026	-0.0029	-0.0029	0.0819	0.0798	0.0798	0.0791	0.0789	0.0786
200	Uniform	Off by 0.5	80	-0.0036	-0.0040	-0.0031	-0.0033	-0.0035	-0.0033	0.0823	0.0787	0.0783	0.0780	0.0782	0.0781
200	Uniform	Off by 1.0	10	-0.0202	-0.0191	-0.0202	-0.0196	-0.0203	-0.0201	0.1279	0.1226	0.1209	0.1207	0.1204	0.1204
200	Uniform	Off by 1.0	20	-0.0169	-0.0186	-0.0182	-0.0182	-0.0182	-0.0184	0.1026	0.0991	0.0980	0.0974	0.0981	0.0969
200	Uniform	Off by 1.0	40	-0.0152	-0.0151	-0.0160	-0.0154	-0.0160	-0.0157	0.0899	0.0877	0.0871	0.0867	0.0859	0.0861
200	Uniform	Off by 1.0	60	-0.0137	-0.0133	-0.0122	-0.0131	-0.0132	-0.0128	0.0840	0.0827	0.0819	0.0811	0.0813	0.0810
200	Uniform	Off by 1.0	80	-0.0137	-0.0138	-0.0131	-0.0134	-0.0133	-0.0133	0.0823	0.0806	0.0801	0.0799	0.0796	0.0798
200	Bimodal	On-target	10	-0.0048	-0.0033	-0.0038	-0.0036	-0.0036	-0.0038	0.1130	0.1065	0.1057	0.1031	0.1029	0.1031
200	Bimodal	On-target	20	0.0019	0.0043	0.0041	0.0035	0.0033	0.0032	0.0948	0.0913	0.0878	0.0885	0.0886	0.0880
200	Bimodal	On-target	40	0.0031	0.0015	0.0021	0.0014	0.0019	0.0020	0.0862	0.0843	0.0827	0.0832	0.0829	0.0826
200	Bimodal	On-target	60	0.0023	0.0033	0.0033	0.0036	0.0035	0.0037	0.0835	0.0824	0.0820	0.0816	0.0816	0.0817
200	Bimodal	On-target	80	0.0039	0.0039	0.0038	0.0038	0.0039	0.0038	0.0803	0.0781	0.0779	0.0778	0.0774	0.0775
200	Bimodal	Off by 0.5	10	-0.0124	-0.0149	-0.0140	-0.0148	-0.0146	-0.0147	0.1147	0.1112	0.1084	0.1080	0.1073	0.1081
200	Bimodal	Off by 0.5	20	-0.0057	-0.0061	-0.0071	-0.0069	-0.0074	-0.0069	0.0974	0.0921	0.0912	0.0900	0.0895	0.0898

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 0.5	40	-0.0065	-0.0078	-0.0075	-0.0072	-0.0072	-0.0074	0.0887	0.0854	0.0851	0.0847	0.0842	0.0845
200	Bimodal	Off by 0.5	60	-0.0047	-0.0051	-0.0055	-0.0057	-0.0056	-0.0056	0.0856	0.0827	0.0827	0.0824	0.0820	0.0821
200	Bimodal	Off by 0.5	80	-0.0025	-0.0026	-0.0028	-0.0028	-0.0029	-0.0030	0.0808	0.0793	0.0785	0.0783	0.0779	0.0780
200	Bimodal	Off by 1.0	10	-0.0271	-0.0291	-0.0298	-0.0292	-0.0287	-0.0292	0.1256	0.1200	0.1191	0.1179	0.1175	0.1173
200	Bimodal	Off by 1.0	20	-0.0181	-0.0194	-0.0196	-0.0199	-0.0191	-0.0191	0.1018	0.0981	0.0972	0.0963	0.0961	0.0965
200	Bimodal	Off by 1.0	40	-0.0189	-0.0174	-0.0185	-0.0181	-0.0178	-0.0179	0.0941	0.0907	0.0906	0.0905	0.0900	0.0899
200	Bimodal	Off by 1.0	60	-0.0156	-0.0163	-0.0157	-0.0160	-0.0162	-0.0157	0.0881	0.0863	0.0859	0.0855	0.0856	0.0855
200	Bimodal	Off by 1.0	80	-0.0131	-0.0133	-0.0134	-0.0134	-0.0138	-0.0135	0.0832	0.0821	0.0814	0.0810	0.0814	0.0813
200	Chisq(5)	On-target	10	0.0038	0.0027	0.0004	0.0010	0.0017	0.0015	0.1120	0.1063	0.1044	0.1034	0.1028	0.1027
200	Chisq(5)	On-target	20	-0.0037	-0.0019	-0.0021	-0.0025	-0.0018	-0.0018	0.0927	0.0880	0.0878	0.0861	0.0853	0.0858
200	Chisq(5)	On-target	40	-0.0039	-0.0041	-0.0042	-0.0037	-0.0039	-0.0038	0.0859	0.0819	0.0818	0.0820	0.0820	0.0815
200	Chisq(5)	On-target	60	-0.0027	-0.0020	-0.0014	-0.0021	-0.0026	-0.0020	0.0816	0.0796	0.0786	0.0787	0.0790	0.0786
200	Chisq(5)	On-target	80	-0.0040	-0.0040	-0.0045	-0.0033	-0.0041	-0.0042	0.0816	0.0797	0.0781	0.0784	0.0784	0.0782
200	Chisq(5)	Off by 0.5	10	-0.0058	-0.0047	-0.0051	-0.0058	-0.0052	-0.0052	0.1123	0.1089	0.1059	0.1060	0.1050	0.1048
200	Chisq(5)	Off by 0.5	20	-0.0074	-0.0062	-0.0075	-0.0074	-0.0069	-0.0069	0.0916	0.0887	0.0878	0.0871	0.0863	0.0867
200	Chisq(5)	Off by 0.5	40	-0.0083	-0.0091	-0.0089	-0.0092	-0.0095	-0.0095	0.0864	0.0835	0.0820	0.0819	0.0819	0.0820
200	Chisq(5)	Off by 0.5	60	-0.0077	-0.0083	-0.0081	-0.0080	-0.0083	-0.0082	0.0814	0.0797	0.0799	0.0793	0.0792	0.0790
200	Chisq(5)	Off by 0.5	80	-0.0099	-0.0102	-0.0102	-0.0097	-0.0100	-0.0099	0.0821	0.0802	0.0801	0.0798	0.0798	0.0798
200	Chisq(5)	Off by 1.0	10	-0.0134	-0.0141	-0.0147	-0.0144	-0.0137	-0.0137	0.1227	0.1166	0.1161	0.1137	0.1132	0.1142
200	Chisq(5)	Off by 1.0	20	-0.0174	-0.0169	-0.0164	-0.0164	-0.0165	-0.0166	0.1005	0.0959	0.0958	0.0948	0.0939	0.0944
200	Chisq(5)	Off by 1.0	40	-0.0177	-0.0178	-0.0187	-0.0182	-0.0184	-0.0180	0.0890	0.0873	0.0872	0.0862	0.0864	0.0861
200	Chisq(5)	Off by 1.0	60	-0.0162	-0.0175	-0.0164	-0.0172	-0.0174	-0.0170	0.0868	0.0845	0.0836	0.0842	0.0837	0.0837
200	Chisq(5)	Off by 1.0	80	-0.0167	-0.0182	-0.0175	-0.0174	-0.0179	-0.0178	0.0854	0.0841	0.0835	0.0834	0.0834	0.0833

Table 2 Bias and RMSE for Mean by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0013	0.0018	0.0017	0.0014	0.0012	0.0014	0.0338	0.0317	0.0317	0.0313	0.0311	0.0311
2000	Normal	On-target	20	0.0002	0.0000	0.0002	-0.0001	0.0002	0.0001	0.0290	0.0278	0.0276	0.0275	0.0274	0.0273
2000	Normal	On-target	40	0.0007	0.0003	0.0004	0.0004	0.0003	0.0005	0.0265	0.0257	0.0253	0.0252	0.0251	0.0252
2000	Normal	On-target	60	0.0002	0.0005	0.0002	0.0003	0.0002	0.0001	0.0245	0.0240	0.0236	0.0235	0.0235	0.0235
2000	Normal	On-target	80	0.0006	0.0008	0.0006	0.0006	0.0006	0.0006	0.0236	0.0231	0.0230	0.0230	0.0229	0.0228
2000	Normal	Off by 0.5	10	0.0026	0.0024	0.0020	0.0019	0.0020	0.0019	0.0345	0.0324	0.0324	0.0318	0.0315	0.0315
2000	Normal	Off by 0.5	20	-0.0003	-0.0008	-0.0008	-0.0007	-0.0008	-0.0008	0.0302	0.0291	0.0290	0.0288	0.0285	0.0285
2000	Normal	Off by 0.5	40	-0.0003	-0.0008	-0.0006	-0.0006	-0.0006	-0.0006	0.0259	0.0251	0.0246	0.0246	0.0247	0.0246
2000	Normal	Off by 0.5	60	-0.0001	-0.0001	0.0002	-0.0001	-0.0002	-0.0002	0.0244	0.0241	0.0238	0.0237	0.0237	0.0236
2000	Normal	Off by 0.5	80	0.0004	0.0003	0.0001	0.0003	0.0002	0.0002	0.0248	0.0243	0.0241	0.0240	0.0240	0.0240
2000	Normal	Off by 1.0	10	-0.0022	-0.0015	-0.0014	-0.0019	-0.0020	-0.0018	0.0374	0.0354	0.0355	0.0352	0.0350	0.0348
2000	Normal	Off by 1.0	20	-0.0015	-0.0017	-0.0014	-0.0014	-0.0016	-0.0015	0.0312	0.0300	0.0296	0.0293	0.0293	0.0292
2000	Normal	Off by 1.0	40	-0.0024	-0.0027	-0.0026	-0.0025	-0.0025	-0.0026	0.0268	0.0258	0.0257	0.0254	0.0255	0.0254
2000	Normal	Off by 1.0	60	-0.0010	-0.0008	-0.0009	-0.0010	-0.0011	-0.0011	0.0258	0.0250	0.0247	0.0246	0.0245	0.0244
2000	Normal	Off by 1.0	80	-0.0037	-0.0038	-0.0037	-0.0038	-0.0037	-0.0038	0.0256	0.0253	0.0251	0.0251	0.0249	0.0250
2000	Normal	Off by 2.5	10	-0.0078	-0.0077	-0.0076	-0.0074	-0.0077	-0.0077	0.0554	0.0538	0.0533	0.0531	0.0533	0.0533
2000	Normal	Off by 2.5	20	-0.0069	-0.0058	-0.0058	-0.0057	-0.0058	-0.0057	0.0411	0.0401	0.0396	0.0390	0.0392	0.0391
2000	Normal	Off by 2.5	40	-0.0094	-0.0085	-0.0087	-0.0087	-0.0088	-0.0088	0.0351	0.0339	0.0339	0.0338	0.0337	0.0338
2000	Normal	Off by 2.5	60	-0.0068	-0.0065	-0.0063	-0.0067	-0.0065	-0.0066	0.0311	0.0300	0.0296	0.0296	0.0296	0.0297
2000	Normal	Off by 2.5	80	-0.0071	-0.0067	-0.0070	-0.0069	-0.0068	-0.0069	0.0286	0.0281	0.0280	0.0279	0.0279	0.0278
2000	Uniform	On-target	10	-0.0005	-0.0011	-0.0008	-0.0006	-0.0007	-0.0008	0.0351	0.0329	0.0325	0.0323	0.0321	0.0321
2000	Uniform	On-target	20	0.0006	0.0002	0.0004	0.0004	0.0003	0.0004	0.0305	0.0291	0.0287	0.0283	0.0282	0.0282
2000	Uniform	On-target	40	0.0007	0.0008	0.0007	0.0008	0.0008	0.0008	0.0258	0.0249	0.0251	0.0247	0.0247	0.0246
2000	Uniform	On-target	60	0.0006	0.0003	0.0004	0.0004	0.0004	0.0004	0.0265	0.0261	0.0258	0.0256	0.0255	0.0256
2000	Uniform	On-target	80	0.0002	0.0003	0.0002	0.0003	0.0002	0.0003	0.0241	0.0237	0.0231	0.0233	0.0234	0.0233
2000	Uniform	Off by 0.5	10	-0.0046	-0.0048	-0.0048	-0.0050	-0.0051	-0.0048	0.0355	0.0346	0.0339	0.0338	0.0340	0.0338
2000	Uniform	Off by 0.5	20	-0.0037	-0.0042	-0.0039	-0.0041	-0.0040	-0.0039	0.0308	0.0290	0.0285	0.0282	0.0281	0.0280
2000	Uniform	Off by 0.5	40	-0.0026	-0.0022	-0.0021	-0.0024	-0.0023	-0.0024	0.0264	0.0256	0.0253	0.0251	0.0252	0.0250
2000	Uniform	Off by 0.5	60	-0.0011	-0.0011	-0.0011	-0.0010	-0.0010	-0.0011	0.0251	0.0244	0.0243	0.0243	0.0242	0.0241
2000	Uniform	Off by 0.5	80	-0.0033	-0.0035	-0.0035	-0.0034	-0.0035	-0.0035	0.0248	0.0245	0.0245	0.0243	0.0244	0.0243
2000	Uniform	Off by 1.0	10	-0.0093	-0.0085	-0.0083	-0.0087	-0.0086	-0.0087	0.0377	0.0366	0.0359	0.0359	0.0356	0.0357
2000	Uniform	Off by 1.0	20	-0.0079	-0.0069	-0.0072	-0.0071	-0.0071	-0.0072	0.0322	0.0305	0.0304	0.0301	0.0299	0.0299
2000	Uniform	Off by 1.0	40	-0.0052	-0.0050	-0.0053	-0.0053	-0.0051	-0.0052	0.0280	0.0271	0.0269	0.0265	0.0266	0.0266
2000	Uniform	Off by 1.0	60	-0.0043	-0.0039	-0.0043	-0.0042	-0.0043	-0.0043	0.0261	0.0253	0.0252	0.0248	0.0250	0.0250
2000	Uniform	Off by 1.0	80	-0.0053	-0.0052	-0.0052	-0.0052	-0.0052	-0.0052	0.0256	0.0248	0.0246	0.0246	0.0245	0.0245
2000	Uniform	Off by 2.5	10	-0.0200	-0.0200	-0.0200	-0.0197	-0.0196	-0.0200	0.0594	0.0586	0.0582	0.0576	0.0575	0.0577
2000	Uniform	Off by 2.5	20	-0.0224	-0.0227	-0.0225	-0.0227	-0.0228	-0.0229	0.0470	0.0472	0.0468	0.0466	0.0463	0.0465
2000	Uniform	Off by 2.5	40	-0.0165	-0.0173	-0.0169	-0.0168	-0.0169	-0.0168	0.0371	0.0372	0.0366	0.0365	0.0364	0.0364
2000	Uniform	Off by 2.5	60	-0.0171	-0.0169	-0.0169	-0.0171	-0.0171	-0.0171	0.0347	0.0339	0.0336	0.0338	0.0338	0.0337
2000	Uniform	Off by 2.5	80	-0.0195	-0.0192	-0.0192	-0.0191	-0.0193	-0.0193	0.0343	0.0334	0.0334	0.0331	0.0332	0.0332
2000	Bimodal	On-target	10	0.0007	-0.0002	0.0001	0.0001	-0.0002	-0.0001	0.0369	0.0343	0.0333	0.0333	0.0330	0.0331
2000	Bimodal	On-target	20	0.0004	0.0000	0.0005	0.0003	0.0004	0.0004	0.0311	0.0296	0.0293	0.0291	0.0288	0.0289

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	0.0012	0.0015	0.0012	0.0013	0.0015	0.0013	0.0292	0.0274	0.0274	0.0272	0.0271	0.0272
2000	Bimodal	On-target	60	-0.0013	-0.0013	-0.0016	-0.0016	-0.0015	-0.0016	0.0268	0.0262	0.0261	0.0258	0.0257	0.0259
2000	Bimodal	On-target	80	-0.0004	-0.0001	-0.0003	-0.0002	-0.0003	-0.0002	0.0246	0.0239	0.0239	0.0237	0.0238	0.0238
2000	Bimodal	Off by 0.5	10	-0.0021	-0.0021	-0.0020	-0.0020	-0.0024	-0.0023	0.0343	0.0324	0.0319	0.0314	0.0313	0.0314
2000	Bimodal	Off by 0.5	20	-0.0028	-0.0028	-0.0031	-0.0031	-0.0032	-0.0031	0.0297	0.0278	0.0274	0.0272	0.0271	0.0269
2000	Bimodal	Off by 0.5	40	-0.0023	-0.0026	-0.0023	-0.0026	-0.0025	-0.0025	0.0253	0.0243	0.0244	0.0240	0.0240	0.0241
2000	Bimodal	Off by 0.5	60	-0.0019	-0.0020	-0.0022	-0.0023	-0.0023	-0.0022	0.0248	0.0242	0.0237	0.0238	0.0237	0.0237
2000	Bimodal	Off by 0.5	80	-0.0020	-0.0017	-0.0019	-0.0018	-0.0018	-0.0017	0.0228	0.0218	0.0218	0.0216	0.0216	0.0216
2000	Bimodal	Off by 1.0	10	-0.0095	-0.0092	-0.0092	-0.0093	-0.0096	-0.0094	0.0394	0.0375	0.0366	0.0365	0.0367	0.0363
2000	Bimodal	Off by 1.0	20	-0.0085	-0.0080	-0.0082	-0.0082	-0.0081	-0.0082	0.0330	0.0314	0.0320	0.0316	0.0315	0.0312
2000	Bimodal	Off by 1.0	40	-0.0072	-0.0073	-0.0067	-0.0071	-0.0071	-0.0070	0.0293	0.0281	0.0278	0.0278	0.0276	0.0277
2000	Bimodal	Off by 1.0	60	-0.0044	-0.0052	-0.0051	-0.0049	-0.0050	-0.0050	0.0279	0.0273	0.0273	0.0271	0.0271	0.0272
2000	Bimodal	Off by 1.0	80	-0.0044	-0.0047	-0.0048	-0.0047	-0.0048	-0.0047	0.0278	0.0272	0.0269	0.0268	0.0271	0.0270
2000	Bimodal	Off by 2.5	10	-0.0192	-0.0192	-0.0191	-0.0192	-0.0194	-0.0196	0.0594	0.0589	0.0582	0.0578	0.0581	0.0582
2000	Bimodal	Off by 2.5	20	-0.0225	-0.0215	-0.0212	-0.0217	-0.0217	-0.0216	0.0481	0.0468	0.0465	0.0462	0.0465	0.0464
2000	Bimodal	Off by 2.5	40	-0.0187	-0.0188	-0.0188	-0.0188	-0.0189	-0.0187	0.0398	0.0388	0.0389	0.0387	0.0386	0.0386
2000	Bimodal	Off by 2.5	60	-0.0195	-0.0197	-0.0198	-0.0195	-0.0197	-0.0197	0.0377	0.0371	0.0371	0.0367	0.0369	0.0367
2000	Bimodal	Off by 2.5	80	-0.0157	-0.0158	-0.0160	-0.0158	-0.0159	-0.0157	0.0348	0.0342	0.0341	0.0339	0.0340	0.0338
2000	Chisq(5)	On-target	10	0.0016	0.0022	0.0023	0.0021	0.0020	0.0021	0.0353	0.0319	0.0315	0.0309	0.0309	0.0311
2000	Chisq(5)	On-target	20	-0.0034	-0.0037	-0.0038	-0.0039	-0.0037	-0.0036	0.0294	0.0283	0.0276	0.0274	0.0274	0.0272
2000	Chisq(5)	On-target	40	-0.0044	-0.0046	-0.0049	-0.0045	-0.0046	-0.0046	0.0271	0.0260	0.0259	0.0260	0.0257	0.0258
2000	Chisq(5)	On-target	60	-0.0043	-0.0044	-0.0043	-0.0045	-0.0044	-0.0044	0.0243	0.0239	0.0237	0.0234	0.0233	0.0232
2000	Chisq(5)	On-target	80	-0.0043	-0.0049	-0.0047	-0.0047	-0.0047	-0.0046	0.0241	0.0240	0.0238	0.0238	0.0237	0.0236
2000	Chisq(5)	Off by 0.5	10	0.0018	0.0025	0.0027	0.0023	0.0024	0.0023	0.0342	0.0330	0.0323	0.0321	0.0317	0.0317
2000	Chisq(5)	Off by 0.5	20	-0.0037	-0.0037	-0.0036	-0.0037	-0.0036	-0.0035	0.0291	0.0278	0.0271	0.0269	0.0267	0.0269
2000	Chisq(5)	Off by 0.5	40	-0.0046	-0.0045	-0.0043	-0.0044	-0.0043	-0.0043	0.0263	0.0256	0.0250	0.0249	0.0249	0.0249
2000	Chisq(5)	Off by 0.5	60	-0.0047	-0.0050	-0.0051	-0.0050	-0.0050	-0.0051	0.0266	0.0260	0.0260	0.0259	0.0258	0.0258
2000	Chisq(5)	Off by 0.5	80	-0.0053	-0.0053	-0.0054	-0.0054	-0.0053	-0.0053	0.0243	0.0242	0.0241	0.0239	0.0239	0.0239
2000	Chisq(5)	Off by 1.0	10	0.0031	0.0040	0.0038	0.0036	0.0036	0.0039	0.0356	0.0336	0.0335	0.0330	0.0327	0.0327
2000	Chisq(5)	Off by 1.0	20	-0.0020	-0.0025	-0.0025	-0.0022	-0.0025	-0.0023	0.0305	0.0286	0.0285	0.0279	0.0281	0.0280
2000	Chisq(5)	Off by 1.0	40	-0.0030	-0.0031	-0.0030	-0.0031	-0.0031	-0.0031	0.0273	0.0261	0.0257	0.0256	0.0255	0.0255
2000	Chisq(5)	Off by 1.0	60	-0.0044	-0.0043	-0.0043	-0.0043	-0.0043	-0.0043	0.0253	0.0253	0.0249	0.0249	0.0247	0.0247
2000	Chisq(5)	Off by 1.0	80	-0.0054	-0.0052	-0.0053	-0.0053	-0.0054	-0.0053	0.0250	0.0242	0.0241	0.0240	0.0240	0.0240
2000	Chisq(5)	Off by 2.5	10	-0.0270	-0.0258	-0.0260	-0.0261	-0.0261	-0.0263	0.0586	0.0566	0.0561	0.0560	0.0559	0.0560
2000	Chisq(5)	Off by 2.5	20	-0.0216	-0.0221	-0.0218	-0.0222	-0.0223	-0.0221	0.0460	0.0445	0.0440	0.0439	0.0439	0.0437
2000	Chisq(5)	Off by 2.5	40	-0.0193	-0.0190	-0.0192	-0.0190	-0.0192	-0.0191	0.0368	0.0356	0.0357	0.0352	0.0353	0.0352
2000	Chisq(5)	Off by 2.5	60	-0.0159	-0.0163	-0.0161	-0.0163	-0.0164	-0.0164	0.0330	0.0319	0.0319	0.0319	0.0318	0.0319
2000	Chisq(5)	Off by 2.5	80	-0.0171	-0.0171	-0.0170	-0.0171	-0.0170	-0.0171	0.0333	0.0325	0.0324	0.0323	0.0324	0.0323

Table 3 Bias and RMSE for Mean by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	0.0000	-0.0001	0.0001	0.0000	-0.0001	-0.0001	0.0147	0.0140	0.0139	0.0140	0.0138	0.0137
10000	Normal	On-target	20	-0.0002	-0.0002	-0.0002	-0.0001	-0.0002	-0.0001	0.0128	0.0123	0.0121	0.0120	0.0119	0.0119
10000	Normal	On-target	40	0.0002	0.0003	0.0001	0.0002	0.0002	0.0002	0.0113	0.0110	0.0108	0.0108	0.0107	0.0108
10000	Normal	On-target	60	-0.0002	-0.0002	0.0000	-0.0002	-0.0002	-0.0001	0.0107	0.0106	0.0104	0.0104	0.0103	0.0104
10000	Normal	On-target	80	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	0.0103	0.0101	0.0101	0.0100	0.0100	0.0100
10000	Normal	Off by 0.5	10	-0.0005	-0.0003	-0.0003	-0.0004	-0.0003	-0.0003	0.0153	0.0141	0.0142	0.0140	0.0140	0.0140
10000	Normal	Off by 0.5	20	-0.0003	-0.0002	0.0000	-0.0001	-0.0002	-0.0002	0.0130	0.0124	0.0122	0.0121	0.0121	0.0121
10000	Normal	Off by 0.5	40	-0.0001	0.0000	-0.0001	0.0001	0.0000	0.0000	0.0117	0.0113	0.0111	0.0111	0.0110	0.0110
10000	Normal	Off by 0.5	60	-0.0002	-0.0004	-0.0004	-0.0004	-0.0003	-0.0004	0.0109	0.0105	0.0105	0.0105	0.0104	0.0104
10000	Normal	Off by 0.5	80	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	0.0104	0.0103	0.0102	0.0102	0.0102	0.0102
10000	Normal	Off by 1.0	10	-0.0007	-0.0004	-0.0004	-0.0004	-0.0005	-0.0005	0.0163	0.0157	0.0156	0.0154	0.0152	0.0153
10000	Normal	Off by 1.0	20	-0.0003	-0.0004	-0.0006	-0.0005	-0.0005	-0.0005	0.0136	0.0129	0.0126	0.0125	0.0125	0.0124
10000	Normal	Off by 1.0	40	0.0000	0.0000	-0.0001	-0.0001	-0.0001	-0.0001	0.0120	0.0115	0.0115	0.0114	0.0114	0.0114
10000	Normal	Off by 1.0	60	-0.0005	-0.0006	-0.0005	-0.0006	-0.0005	-0.0005	0.0109	0.0106	0.0105	0.0105	0.0106	0.0106
10000	Normal	Off by 1.0	80	-0.0012	-0.0011	-0.0010	-0.0011	-0.0011	-0.0011	0.0108	0.0104	0.0103	0.0102	0.0102	0.0102
10000	Normal	Off by 2.5	10	-0.0022	-0.0026	-0.0027	-0.0027	-0.0025	-0.0026	0.0252	0.0247	0.0245	0.0246	0.0245	0.0244
10000	Normal	Off by 2.5	20	-0.0014	-0.0013	-0.0015	-0.0013	-0.0014	-0.0014	0.0175	0.0171	0.0169	0.0168	0.0167	0.0166
10000	Normal	Off by 2.5	40	-0.0018	-0.0019	-0.0019	-0.0018	-0.0019	-0.0019	0.0146	0.0142	0.0140	0.0140	0.0140	0.0139
10000	Normal	Off by 2.5	60	-0.0017	-0.0017	-0.0016	-0.0016	-0.0016	-0.0016	0.0129	0.0123	0.0121	0.0122	0.0121	0.0122
10000	Normal	Off by 2.5	80	-0.0023	-0.0023	-0.0022	-0.0023	-0.0023	-0.0023	0.0124	0.0120	0.0120	0.0119	0.0119	0.0119
10000	Uniform	On-target	10	0.0002	0.0002	0.0002	0.0001	0.0000	0.0001	0.0150	0.0143	0.0141	0.0138	0.0139	0.0138
10000	Uniform	On-target	20	-0.0002	0.0000	-0.0001	0.0000	-0.0001	-0.0001	0.0128	0.0122	0.0119	0.0119	0.0119	0.0119
10000	Uniform	On-target	40	0.0002	0.0001	0.0000	0.0001	0.0001	0.0001	0.0119	0.0112	0.0110	0.0110	0.0110	0.0109
10000	Uniform	On-target	60	0.0000	-0.0001	-0.0002	0.0000	-0.0001	-0.0001	0.0108	0.0104	0.0105	0.0104	0.0104	0.0104
10000	Uniform	On-target	80	-0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0102	0.0103	0.0102	0.0102	0.0102
10000	Uniform	Off by 0.5	10	-0.0037	-0.0035	-0.0034	-0.0035	-0.0035	-0.0035	0.0161	0.0153	0.0148	0.0147	0.0147	0.0146
10000	Uniform	Off by 0.5	20	-0.0025	-0.0028	-0.0027	-0.0028	-0.0028	-0.0027	0.0132	0.0126	0.0124	0.0122	0.0122	0.0122
10000	Uniform	Off by 0.5	40	-0.0019	-0.0019	-0.0019	-0.0020	-0.0020	-0.0019	0.0123	0.0115	0.0114	0.0113	0.0113	0.0112
10000	Uniform	Off by 0.5	60	-0.0018	-0.0019	-0.0019	-0.0018	-0.0018	-0.0018	0.0112	0.0107	0.0107	0.0107	0.0106	0.0106
10000	Uniform	Off by 0.5	80	-0.0022	-0.0021	-0.0022	-0.0022	-0.0022	-0.0022	0.0109	0.0108	0.0107	0.0107	0.0107	0.0107
10000	Uniform	Off by 1.0	10	-0.0080	-0.0075	-0.0078	-0.0078	-0.0077	-0.0077	0.0178	0.0172	0.0171	0.0171	0.0170	0.0170
10000	Uniform	Off by 1.0	20	-0.0056	-0.0055	-0.0055	-0.0055	-0.0056	-0.0055	0.0143	0.0138	0.0139	0.0136	0.0136	0.0135
10000	Uniform	Off by 1.0	40	-0.0042	-0.0042	-0.0042	-0.0041	-0.0042	-0.0041	0.0129	0.0123	0.0123	0.0121	0.0121	0.0121
10000	Uniform	Off by 1.0	60	-0.0039	-0.0039	-0.0038	-0.0038	-0.0039	-0.0038	0.0116	0.0112	0.0111	0.0112	0.0112	0.0112
10000	Uniform	Off by 1.0	80	-0.0044	-0.0045	-0.0045	-0.0045	-0.0045	-0.0045	0.0117	0.0116	0.0115	0.0115	0.0114	0.0114
10000	Uniform	Off by 2.5	10	-0.0135	-0.0134	-0.0134	-0.0135	-0.0135	-0.0135	0.0278	0.0272	0.0269	0.0269	0.0268	0.0269
10000	Uniform	Off by 2.5	20	-0.0125	-0.0127	-0.0129	-0.0127	-0.0129	-0.0128	0.0218	0.0215	0.0216	0.0213	0.0214	0.0212
10000	Uniform	Off by 2.5	40	-0.0120	-0.0121	-0.0121	-0.0121	-0.0121	-0.0121	0.0194	0.0190	0.0190	0.0190	0.0189	0.0190
10000	Uniform	Off by 2.5	60	-0.0119	-0.0117	-0.0117	-0.0118	-0.0117	-0.0118	0.0178	0.0174	0.0172	0.0173	0.0173	0.0173
10000	Uniform	Off by 2.5	80	-0.0117	-0.0118	-0.0118	-0.0118	-0.0118	-0.0118	0.0172	0.0171	0.0169	0.0169	0.0169	0.0169
10000	Bimodal	On-target	10	0.0003	0.0004	0.0003	0.0002	0.0003	0.0003	0.0160	0.0152	0.0148	0.0148	0.0147	0.0147
10000	Bimodal	On-target	20	-0.0001	0.0001	0.0002	-0.0001	0.0000	0.0001	0.0138	0.0129	0.0130	0.0127	0.0128	0.0127

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	0.0003	0.0001	0.0001	0.0002	0.0002	0.0001	0.0128	0.0122	0.0121	0.0121	0.0121	0.0120
10000	Bimodal	On-target	60	-0.0006	-0.0005	-0.0004	-0.0005	-0.0004	-0.0004	0.0124	0.0119	0.0121	0.0119	0.0119	0.0118
10000	Bimodal	On-target	80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0113	0.0111	0.0111	0.0110	0.0111	0.0110
10000	Bimodal	Off by 0.5	10	-0.0036	-0.0038	-0.0039	-0.0038	-0.0038	-0.0037	0.0163	0.0157	0.0154	0.0153	0.0153	0.0152
10000	Bimodal	Off by 0.5	20	-0.0029	-0.0028	-0.0028	-0.0028	-0.0028	-0.0028	0.0146	0.0137	0.0135	0.0133	0.0133	0.0133
10000	Bimodal	Off by 0.5	40	-0.0021	-0.0020	-0.0020	-0.0020	-0.0019	-0.0020	0.0132	0.0125	0.0123	0.0124	0.0123	0.0124
10000	Bimodal	Off by 0.5	60	-0.0017	-0.0017	-0.0017	-0.0017	-0.0017	-0.0017	0.0120	0.0118	0.0119	0.0117	0.0117	0.0117
10000	Bimodal	Off by 0.5	80	-0.0017	-0.0017	-0.0016	-0.0016	-0.0017	-0.0016	0.0116	0.0115	0.0113	0.0113	0.0113	0.0113
10000	Bimodal	Off by 1.0	10	-0.0080	-0.0077	-0.0077	-0.0078	-0.0077	-0.0078	0.0184	0.0176	0.0174	0.0174	0.0173	0.0172
10000	Bimodal	Off by 1.0	20	-0.0060	-0.0061	-0.0059	-0.0060	-0.0059	-0.0060	0.0154	0.0150	0.0149	0.0148	0.0146	0.0147
10000	Bimodal	Off by 1.0	40	-0.0043	-0.0043	-0.0043	-0.0043	-0.0042	-0.0043	0.0138	0.0133	0.0131	0.0132	0.0131	0.0131
10000	Bimodal	Off by 1.0	60	-0.0036	-0.0036	-0.0036	-0.0036	-0.0036	-0.0037	0.0129	0.0127	0.0126	0.0126	0.0126	0.0126
10000	Bimodal	Off by 1.0	80	-0.0035	-0.0034	-0.0033	-0.0034	-0.0034	-0.0034	0.0122	0.0120	0.0119	0.0119	0.0119	0.0119
10000	Bimodal	Off by 2.5	10	-0.0142	-0.0142	-0.0142	-0.0143	-0.0143	-0.0142	0.0283	0.0277	0.0276	0.0276	0.0276	0.0275
10000	Bimodal	Off by 2.5	20	-0.0149	-0.0150	-0.0148	-0.0148	-0.0148	-0.0149	0.0241	0.0235	0.0235	0.0234	0.0234	0.0234
10000	Bimodal	Off by 2.5	40	-0.0140	-0.0140	-0.0138	-0.0138	-0.0140	-0.0139	0.0206	0.0201	0.0198	0.0198	0.0199	0.0199
10000	Bimodal	Off by 2.5	60	-0.0125	-0.0123	-0.0124	-0.0122	-0.0123	-0.0123	0.0188	0.0186	0.0185	0.0184	0.0184	0.0184
10000	Bimodal	Off by 2.5	80	-0.0109	-0.0109	-0.0109	-0.0109	-0.0110	-0.0109	0.0176	0.0173	0.0172	0.0172	0.0172	0.0172
10000	Chisq(5)	On-target	10	0.0020	0.0017	0.0018	0.0019	0.0018	0.0017	0.0157	0.0146	0.0144	0.0143	0.0142	0.0143
10000	Chisq(5)	On-target	20	-0.0023	-0.0024	-0.0022	-0.0023	-0.0024	-0.0023	0.0132	0.0127	0.0125	0.0126	0.0124	0.0123
10000	Chisq(5)	On-target	40	-0.0040	-0.0041	-0.0040	-0.0040	-0.0041	-0.0040	0.0127	0.0124	0.0121	0.0121	0.0120	0.0120
10000	Chisq(5)	On-target	60	-0.0046	-0.0044	-0.0045	-0.0045	-0.0046	-0.0045	0.0122	0.0119	0.0118	0.0118	0.0118	0.0117
10000	Chisq(5)	On-target	80	-0.0047	-0.0048	-0.0048	-0.0047	-0.0047	-0.0047	0.0125	0.0122	0.0122	0.0122	0.0122	0.0122
10000	Chisq(5)	Off by 0.5	10	0.0038	0.0036	0.0038	0.0037	0.0037	0.0036	0.0163	0.0151	0.0151	0.0149	0.0148	0.0149
10000	Chisq(5)	Off by 0.5	20	-0.0010	-0.0010	-0.0009	-0.0009	-0.0009	-0.0009	0.0133	0.0123	0.0123	0.0122	0.0122	0.0122
10000	Chisq(5)	Off by 0.5	40	-0.0028	-0.0028	-0.0028	-0.0028	-0.0027	-0.0027	0.0121	0.0119	0.0118	0.0117	0.0116	0.0116
10000	Chisq(5)	Off by 0.5	60	-0.0038	-0.0037	-0.0037	-0.0037	-0.0037	-0.0036	0.0122	0.0117	0.0116	0.0116	0.0117	0.0116
10000	Chisq(5)	Off by 0.5	80	-0.0044	-0.0043	-0.0043	-0.0043	-0.0043	-0.0043	0.0122	0.0121	0.0121	0.0121	0.0121	0.0121
10000	Chisq(5)	Off by 1.0	10	0.0031	0.0032	0.0031	0.0033	0.0032	0.0033	0.0168	0.0161	0.0158	0.0157	0.0156	0.0156
10000	Chisq(5)	Off by 1.0	20	-0.0003	-0.0004	-0.0004	-0.0005	-0.0005	-0.0004	0.0138	0.0131	0.0129	0.0129	0.0127	0.0127
10000	Chisq(5)	Off by 1.0	40	-0.0025	-0.0026	-0.0024	-0.0026	-0.0025	-0.0026	0.0124	0.0119	0.0119	0.0118	0.0117	0.0118
10000	Chisq(5)	Off by 1.0	60	-0.0036	-0.0034	-0.0035	-0.0036	-0.0036	-0.0036	0.0119	0.0116	0.0116	0.0116	0.0116	0.0116
10000	Chisq(5)	Off by 1.0	80	-0.0044	-0.0045	-0.0044	-0.0044	-0.0044	-0.0044	0.0124	0.0122	0.0122	0.0122	0.0121	0.0121
10000	Chisq(5)	Off by 2.5	10	-0.0208	-0.0209	-0.0208	-0.0208	-0.0209	-0.0208	0.0311	0.0306	0.0305	0.0304	0.0304	0.0303
10000	Chisq(5)	Off by 2.5	20	-0.0176	-0.0176	-0.0173	-0.0175	-0.0175	-0.0175	0.0247	0.0241	0.0238	0.0240	0.0240	0.0239
10000	Chisq(5)	Off by 2.5	40	-0.0140	-0.0141	-0.0141	-0.0141	-0.0139	-0.0141	0.0208	0.0204	0.0203	0.0203	0.0202	0.0202
10000	Chisq(5)	Off by 2.5	60	-0.0126	-0.0123	-0.0124	-0.0123	-0.0124	-0.0124	0.0185	0.0181	0.0180	0.0179	0.0180	0.0179
10000	Chisq(5)	Off by 2.5	80	-0.0113	-0.0113	-0.0113	-0.0113	-0.0113	-0.0113	0.0169	0.0166	0.0166	0.0166	0.0166	0.0166

Table 4 Bias and RMSE for Variance by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0195	0.0185	0.0203	0.0186	0.0175	0.0180	0.2308	0.2221	0.2212	0.2178	0.2173	0.2162
200	Normal	On-target	20	0.0180	0.0170	0.0166	0.0172	0.0158	0.0169	0.1723	0.1617	0.1576	0.1563	0.1566	0.1555
200	Normal	On-target	40	0.0181	0.0164	0.0144	0.0165	0.0159	0.0167	0.1423	0.1316	0.1298	0.1301	0.1301	0.1297
200	Normal	On-target	60	0.0112	0.0131	0.0146	0.0132	0.0143	0.0134	0.1308	0.1243	0.1234	0.1234	0.1226	0.1224
200	Normal	On-target	80	0.0116	0.0126	0.0113	0.0121	0.0129	0.0130	0.1246	0.1187	0.1182	0.1172	0.1170	0.1174
200	Normal	Off by 0.5	10	0.0190	0.0177	0.0192	0.0175	0.0194	0.0186	0.2354	0.2240	0.2229	0.2208	0.2222	0.2209
200	Normal	Off by 0.5	20	0.0199	0.0191	0.0180	0.0180	0.0177	0.0185	0.1733	0.1644	0.1622	0.1584	0.1590	0.1586
200	Normal	Off by 0.5	40	0.0179	0.0155	0.0139	0.0152	0.0139	0.0147	0.1408	0.1333	0.1303	0.1304	0.1293	0.1287
200	Normal	Off by 0.5	60	0.0146	0.0150	0.0137	0.0139	0.0133	0.0136	0.1308	0.1253	0.1243	0.1236	0.1231	0.1233
200	Normal	Off by 0.5	80	0.0125	0.0117	0.0125	0.0122	0.0121	0.0128	0.1231	0.1162	0.1172	0.1154	0.1152	0.1149
200	Normal	Off by 1.0	10	0.0101	0.0089	0.0112	0.0115	0.0114	0.0108	0.2365	0.2224	0.2195	0.2205	0.2190	0.2192
200	Normal	Off by 1.0	20	0.0223	0.0224	0.0230	0.0222	0.0239	0.0240	0.1761	0.1677	0.1675	0.1637	0.1644	0.1644
200	Normal	Off by 1.0	40	0.0162	0.0139	0.0133	0.0148	0.0148	0.0151	0.1425	0.1352	0.1333	0.1333	0.1328	0.1329
200	Normal	Off by 1.0	60	0.0143	0.0139	0.0139	0.0146	0.0139	0.0144	0.1347	0.1262	0.1245	0.1238	0.1241	0.1241
200	Normal	Off by 1.0	80	0.0105	0.0114	0.0110	0.0099	0.0105	0.0102	0.1226	0.1173	0.1171	0.1164	0.1156	0.1156
200	Uniform	On-target	10	0.0557	0.0567	0.0539	0.0546	0.0536	0.0537	0.2337	0.2258	0.2218	0.2197	0.2191	0.2194
200	Uniform	On-target	20	0.0505	0.0492	0.0500	0.0499	0.0491	0.0497	0.1614	0.1498	0.1506	0.1462	0.1448	0.1450
200	Uniform	On-target	40	0.0353	0.0327	0.0337	0.0333	0.0322	0.0332	0.1171	0.1093	0.1080	0.1057	0.1058	0.1053
200	Uniform	On-target	60	0.0306	0.0316	0.0327	0.0307	0.0318	0.0319	0.1053	0.0999	0.0983	0.0961	0.0965	0.0959
200	Uniform	On-target	80	0.0310	0.0322	0.0326	0.0330	0.0322	0.0327	0.0977	0.0921	0.0916	0.0907	0.0901	0.0900
200	Uniform	Off by 0.5	10	0.0596	0.0625	0.0588	0.0588	0.0599	0.0598	0.2365	0.2256	0.2183	0.2176	0.2172	0.2175
200	Uniform	Off by 0.5	20	0.0549	0.0552	0.0546	0.0525	0.0532	0.0533	0.1617	0.1509	0.1484	0.1461	0.1450	0.1449
200	Uniform	Off by 0.5	40	0.0372	0.0370	0.0368	0.0366	0.0366	0.0367	0.1207	0.1130	0.1097	0.1083	0.1083	0.1080
200	Uniform	Off by 0.5	60	0.0371	0.0343	0.0349	0.0348	0.0354	0.0354	0.1086	0.1005	0.0992	0.0976	0.0977	0.0965
200	Uniform	Off by 0.5	80	0.0328	0.0340	0.0344	0.0347	0.0343	0.0340	0.0998	0.0942	0.0939	0.0925	0.0924	0.0920
200	Uniform	Off by 1.0	10	0.0669	0.0694	0.0691	0.0692	0.0689	0.0688	0.2350	0.2231	0.2233	0.2206	0.2194	0.2188
200	Uniform	Off by 1.0	20	0.0613	0.0598	0.0614	0.0599	0.0605	0.0602	0.1668	0.1583	0.1563	0.1543	0.1539	0.1520
200	Uniform	Off by 1.0	40	0.0398	0.0386	0.0397	0.0405	0.0400	0.0396	0.1224	0.1120	0.1103	0.1085	0.1086	0.1074
200	Uniform	Off by 1.0	60	0.0413	0.0399	0.0401	0.0409	0.0402	0.0394	0.1173	0.1070	0.1043	0.1032	0.1030	0.1021
200	Uniform	Off by 1.0	80	0.0340	0.0337	0.0359	0.0351	0.0354	0.0359	0.1025	0.0973	0.0959	0.0949	0.0951	0.0944
200	Bimodal	On-target	10	0.0438	0.0450	0.0424	0.0438	0.0440	0.0436	0.2062	0.1948	0.1903	0.1913	0.1900	0.1906
200	Bimodal	On-target	20	0.0398	0.0401	0.0420	0.0387	0.0406	0.0425	0.1453	0.1334	0.1303	0.1288	0.1289	0.1293
200	Bimodal	On-target	40	0.0206	0.0259	0.0239	0.0245	0.0245	0.0243	0.0990	0.0917	0.0884	0.0860	0.0857	0.0859
200	Bimodal	On-target	60	0.0247	0.0243	0.0263	0.0256	0.0251	0.0251	0.0865	0.0790	0.0764	0.0747	0.0736	0.0734
200	Bimodal	On-target	80	0.0224	0.0234	0.0226	0.0232	0.0224	0.0227	0.0749	0.0661	0.0629	0.0637	0.0626	0.0625
200	Bimodal	Off by 0.5	10	0.0423	0.0420	0.0431	0.0426	0.0435	0.0437	0.2103	0.1973	0.1945	0.1952	0.1933	0.1932
200	Bimodal	Off by 0.5	20	0.0391	0.0420	0.0397	0.0406	0.0394	0.0399	0.1455	0.1365	0.1342	0.1326	0.1302	0.1300
200	Bimodal	Off by 0.5	40	0.0285	0.0286	0.0290	0.0284	0.0282	0.0286	0.1030	0.0930	0.0892	0.0886	0.0874	0.0871
200	Bimodal	Off by 0.5	60	0.0281	0.0266	0.0281	0.0266	0.0271	0.0271	0.0871	0.0776	0.0761	0.0736	0.0731	0.0732
200	Bimodal	Off by 0.5	80	0.0237	0.0238	0.0227	0.0234	0.0237	0.0231	0.0757	0.0691	0.0655	0.0651	0.0645	0.0640
200	Bimodal	Off by 1.0	10	0.0579	0.0547	0.0563	0.0550	0.0550	0.0550	0.2178	0.2050	0.2078	0.2039	0.2031	0.2016
200	Bimodal	Off by 1.0	20	0.0444	0.0449	0.0465	0.0452	0.0440	0.0450	0.1549	0.1418	0.1404	0.1366	0.1358	0.1361

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 1.0	40	0.0344	0.0344	0.0346	0.0331	0.0331	0.0332	0.1055	0.0942	0.0950	0.0913	0.0905	0.0904
200	Bimodal	Off by 1.0	60	0.0356	0.0328	0.0337	0.0335	0.0332	0.0329	0.0911	0.0806	0.0793	0.0770	0.0776	0.0764
200	Bimodal	Off by 1.0	80	0.0295	0.0292	0.0290	0.0293	0.0299	0.0291	0.0794	0.0716	0.0700	0.0687	0.0684	0.0680
200	Chisq(5)	On-target	10	-0.0380	-0.0382	-0.0411	-0.0407	-0.0400	-0.0394	0.2276	0.2219	0.2197	0.2175	0.2175	0.2179
200	Chisq(5)	On-target	20	-0.0369	-0.0382	-0.0389	-0.0384	-0.0404	-0.0392	0.1799	0.1682	0.1670	0.1670	0.1655	0.1665
200	Chisq(5)	On-target	40	-0.0385	-0.0362	-0.0367	-0.0368	-0.0378	-0.0373	0.1526	0.1487	0.1466	0.1452	0.1452	0.1453
200	Chisq(5)	On-target	60	-0.0361	-0.0360	-0.0354	-0.0358	-0.0357	-0.0356	0.1512	0.1448	0.1446	0.1435	0.1432	0.1433
200	Chisq(5)	On-target	80	-0.0295	-0.0325	-0.0307	-0.0303	-0.0304	-0.0302	0.1434	0.1417	0.1407	0.1402	0.1397	0.1400
200	Chisq(5)	Off by 0.5	10	0.0071	0.0056	0.0071	0.0051	0.0043	0.0053	0.2365	0.2274	0.2278	0.2231	0.2244	0.2239
200	Chisq(5)	Off by 0.5	20	-0.0004	-0.0044	-0.0043	-0.0035	-0.0051	-0.0029	0.1811	0.1699	0.1673	0.1666	0.1645	0.1651
200	Chisq(5)	Off by 0.5	40	-0.0200	-0.0157	-0.0148	-0.0163	-0.0160	-0.0154	0.1554	0.1484	0.1474	0.1460	0.1461	0.1467
200	Chisq(5)	Off by 0.5	60	-0.0169	-0.0169	-0.0176	-0.0157	-0.0169	-0.0162	0.1516	0.1442	0.1455	0.1450	0.1444	0.1434
200	Chisq(5)	Off by 0.5	80	-0.0166	-0.0170	-0.0175	-0.0169	-0.0173	-0.0168	0.1469	0.1418	0.1422	0.1412	0.1412	0.1415
200	Chisq(5)	Off by 1.0	10	0.0497	0.0487	0.0475	0.0498	0.0491	0.0484	0.2616	0.2478	0.2433	0.2435	0.2424	0.2418
200	Chisq(5)	Off by 1.0	20	0.0286	0.0292	0.0301	0.0298	0.0292	0.0295	0.1873	0.1808	0.1784	0.1776	0.1775	0.1777
200	Chisq(5)	Off by 1.0	40	0.0114	0.0112	0.0082	0.0088	0.0088	0.0095	0.1639	0.1557	0.1540	0.1532	0.1531	0.1528
200	Chisq(5)	Off by 1.0	60	0.0020	0.0040	0.0035	0.0041	0.0037	0.0041	0.1565	0.1496	0.1482	0.1470	0.1465	0.1466
200	Chisq(5)	Off by 1.0	80	-0.0066	-0.0044	-0.0053	-0.0046	-0.0048	-0.0049	0.1479	0.1449	0.1448	0.1436	0.1437	0.1440

Table 5 Bias and RMSE for Variance by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0041	0.0056	0.0045	0.0051	0.0046	0.0049	0.0739	0.0695	0.0687	0.0681	0.0679	0.0678
2000	Normal	On-target	20	0.0007	0.0018	0.0023	0.0020	0.0020	0.0020	0.0527	0.0490	0.0485	0.0479	0.0480	0.0477
2000	Normal	On-target	40	0.0000	0.0001	0.0002	0.0002	-0.0001	0.0001	0.0424	0.0407	0.0404	0.0399	0.0397	0.0399
2000	Normal	On-target	60	0.0016	0.0021	0.0023	0.0022	0.0023	0.0022	0.0397	0.0380	0.0374	0.0378	0.0375	0.0373
2000	Normal	On-target	80	0.0022	0.0021	0.0025	0.0023	0.0023	0.0024	0.0389	0.0375	0.0370	0.0370	0.0369	0.0367
2000	Normal	Off by 0.5	10	0.0012	-0.0010	-0.0004	-0.0001	-0.0005	-0.0004	0.0692	0.0659	0.0652	0.0644	0.0645	0.0642
2000	Normal	Off by 0.5	20	0.0014	0.0012	0.0001	0.0001	0.0003	0.0003	0.0528	0.0497	0.0491	0.0485	0.0484	0.0481
2000	Normal	Off by 0.5	40	0.0015	0.0007	0.0009	0.0010	0.0010	0.0011	0.0432	0.0412	0.0409	0.0401	0.0401	0.0399
2000	Normal	Off by 0.5	60	-0.0011	-0.0005	-0.0004	-0.0003	-0.0001	-0.0005	0.0393	0.0379	0.0374	0.0373	0.0369	0.0371
2000	Normal	Off by 0.5	80	0.0048	0.0048	0.0044	0.0045	0.0047	0.0047	0.0400	0.0382	0.0381	0.0380	0.0379	0.0379
2000	Normal	Off by 1.0	10	0.0040	0.0029	0.0034	0.0034	0.0032	0.0030	0.0737	0.0693	0.0690	0.0681	0.0681	0.0678
2000	Normal	Off by 1.0	20	0.0000	0.0000	0.0003	0.0002	0.0003	0.0002	0.0532	0.0493	0.0490	0.0479	0.0479	0.0480
2000	Normal	Off by 1.0	40	0.0014	0.0013	0.0018	0.0018	0.0015	0.0016	0.0440	0.0423	0.0413	0.0409	0.0410	0.0408
2000	Normal	Off by 1.0	60	0.0016	0.0020	0.0017	0.0018	0.0017	0.0017	0.0402	0.0389	0.0380	0.0380	0.0381	0.0379
2000	Normal	Off by 1.0	80	0.0029	0.0031	0.0032	0.0029	0.0031	0.0032	0.0382	0.0374	0.0370	0.0365	0.0368	0.0366
2000	Normal	Off by 2.5	10	0.0027	0.0032	0.0024	0.0031	0.0023	0.0028	0.0845	0.0804	0.0804	0.0795	0.0792	0.0795
2000	Normal	Off by 2.5	20	0.0002	-0.0002	0.0001	-0.0003	-0.0002	-0.0005	0.0616	0.0575	0.0566	0.0557	0.0559	0.0556
2000	Normal	Off by 2.5	40	0.0021	0.0013	0.0011	0.0010	0.0012	0.0012	0.0490	0.0457	0.0450	0.0448	0.0447	0.0446
2000	Normal	Off by 2.5	60	0.0009	-0.0005	-0.0003	-0.0004	-0.0002	0.0000	0.0442	0.0407	0.0408	0.0403	0.0400	0.0399
2000	Normal	Off by 2.5	80	-0.0019	-0.0026	-0.0021	-0.0022	-0.0025	-0.0026	0.0418	0.0381	0.0382	0.0377	0.0374	0.0375
2000	Uniform	On-target	10	0.0281	0.0272	0.0271	0.0280	0.0272	0.0276	0.0725	0.0702	0.0695	0.0697	0.0694	0.0694
2000	Uniform	On-target	20	0.0239	0.0244	0.0240	0.0246	0.0241	0.0244	0.0522	0.0485	0.0480	0.0484	0.0475	0.0473
2000	Uniform	On-target	40	0.0161	0.0152	0.0157	0.0160	0.0159	0.0161	0.0396	0.0371	0.0362	0.0362	0.0361	0.0360
2000	Uniform	On-target	60	0.0148	0.0150	0.0153	0.0151	0.0151	0.0150	0.0355	0.0331	0.0328	0.0326	0.0324	0.0323
2000	Uniform	On-target	80	0.0164	0.0157	0.0162	0.0161	0.0160	0.0160	0.0337	0.0319	0.0319	0.0317	0.0317	0.0316
2000	Uniform	Off by 0.5	10	0.0362	0.0367	0.0361	0.0362	0.0367	0.0359	0.0766	0.0737	0.0735	0.0719	0.0727	0.0718
2000	Uniform	Off by 0.5	20	0.0264	0.0260	0.0263	0.0260	0.0264	0.0262	0.0541	0.0510	0.0501	0.0497	0.0497	0.0496
2000	Uniform	Off by 0.5	40	0.0196	0.0192	0.0194	0.0193	0.0196	0.0194	0.0419	0.0389	0.0384	0.0379	0.0379	0.0376
2000	Uniform	Off by 0.5	60	0.0190	0.0175	0.0181	0.0182	0.0181	0.0179	0.0373	0.0346	0.0345	0.0344	0.0341	0.0340
2000	Uniform	Off by 0.5	80	0.0163	0.0166	0.0157	0.0160	0.0161	0.0161	0.0345	0.0326	0.0317	0.0318	0.0316	0.0316
2000	Uniform	Off by 1.0	10	0.0394	0.0396	0.0398	0.0404	0.0402	0.0398	0.0779	0.0748	0.0739	0.0740	0.0741	0.0740
2000	Uniform	Off by 1.0	20	0.0342	0.0351	0.0346	0.0350	0.0348	0.0349	0.0599	0.0573	0.0559	0.0558	0.0560	0.0558
2000	Uniform	Off by 1.0	40	0.0267	0.0268	0.0270	0.0269	0.0272	0.0270	0.0451	0.0432	0.0425	0.0425	0.0425	0.0423
2000	Uniform	Off by 1.0	60	0.0225	0.0218	0.0222	0.0223	0.0223	0.0223	0.0403	0.0373	0.0371	0.0366	0.0366	0.0366
2000	Uniform	Off by 1.0	80	0.0219	0.0215	0.0215	0.0215	0.0216	0.0215	0.0370	0.0350	0.0349	0.0345	0.0346	0.0344
2000	Uniform	Off by 2.5	10	0.0247	0.0253	0.0244	0.0254	0.0251	0.0252	0.0839	0.0822	0.0804	0.0805	0.0803	0.0801
2000	Uniform	Off by 2.5	20	0.0403	0.0411	0.0407	0.0410	0.0411	0.0414	0.0698	0.0681	0.0667	0.0660	0.0664	0.0662
2000	Uniform	Off by 2.5	40	0.0454	0.0457	0.0459	0.0456	0.0461	0.0455	0.0636	0.0615	0.0610	0.0602	0.0604	0.0601
2000	Uniform	Off by 2.5	60	0.0490	0.0480	0.0482	0.0481	0.0481	0.0482	0.0626	0.0600	0.0599	0.0593	0.0595	0.0593
2000	Uniform	Off by 2.5	80	0.0465	0.0462	0.0468	0.0464	0.0464	0.0463	0.0591	0.0572	0.0573	0.0565	0.0564	0.0562
2000	Bimodal	On-target	10	0.0298	0.0281	0.0285	0.0285	0.0288	0.0286	0.0685	0.0648	0.0628	0.0634	0.0629	0.0624
2000	Bimodal	On-target	20	0.0247	0.0254	0.0254	0.0252	0.0253	0.0254	0.0488	0.0465	0.0458	0.0451	0.0449	0.0445

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	0.0158	0.0148	0.0155	0.0149	0.0155	0.0155	0.0352	0.0310	0.0314	0.0303	0.0303	0.0303
2000	Bimodal	On-target	60	0.0133	0.0134	0.0137	0.0137	0.0140	0.0138	0.0291	0.0270	0.0264	0.0261	0.0260	0.0259
2000	Bimodal	On-target	80	0.0116	0.0123	0.0119	0.0121	0.0120	0.0119	0.0257	0.0238	0.0233	0.0231	0.0227	0.0228
2000	Bimodal	Off by 0.5	10	0.0288	0.0283	0.0288	0.0281	0.0283	0.0276	0.0693	0.0652	0.0647	0.0635	0.0636	0.0629
2000	Bimodal	Off by 0.5	20	0.0287	0.0289	0.0285	0.0291	0.0288	0.0288	0.0526	0.0492	0.0484	0.0483	0.0479	0.0479
2000	Bimodal	Off by 0.5	40	0.0190	0.0197	0.0198	0.0192	0.0195	0.0195	0.0379	0.0343	0.0341	0.0335	0.0334	0.0331
2000	Bimodal	Off by 0.5	60	0.0158	0.0154	0.0150	0.0154	0.0153	0.0153	0.0309	0.0278	0.0270	0.0266	0.0266	0.0265
2000	Bimodal	Off by 0.5	80	0.0152	0.0145	0.0147	0.0143	0.0144	0.0143	0.0279	0.0251	0.0250	0.0244	0.0243	0.0241
2000	Bimodal	Off by 1.0	10	0.0371	0.0387	0.0390	0.0384	0.0387	0.0385	0.0762	0.0738	0.0731	0.0726	0.0725	0.0719
2000	Bimodal	Off by 1.0	20	0.0338	0.0336	0.0332	0.0333	0.0335	0.0336	0.0570	0.0537	0.0527	0.0524	0.0522	0.0522
2000	Bimodal	Off by 1.0	40	0.0252	0.0253	0.0255	0.0258	0.0256	0.0254	0.0405	0.0376	0.0370	0.0370	0.0367	0.0364
2000	Bimodal	Off by 1.0	60	0.0188	0.0193	0.0190	0.0195	0.0197	0.0194	0.0325	0.0298	0.0294	0.0289	0.0291	0.0291
2000	Bimodal	Off by 1.0	80	0.0186	0.0178	0.0180	0.0179	0.0180	0.0180	0.0299	0.0270	0.0270	0.0269	0.0266	0.0267
2000	Bimodal	Off by 2.5	10	0.0293	0.0293	0.0289	0.0291	0.0285	0.0290	0.0826	0.0800	0.0791	0.0782	0.0777	0.0779
2000	Bimodal	Off by 2.5	20	0.0438	0.0436	0.0430	0.0434	0.0435	0.0435	0.0687	0.0659	0.0639	0.0638	0.0639	0.0639
2000	Bimodal	Off by 2.5	40	0.0528	0.0543	0.0535	0.0537	0.0537	0.0538	0.0662	0.0646	0.0633	0.0633	0.0632	0.0632
2000	Bimodal	Off by 2.5	60	0.0499	0.0510	0.0504	0.0509	0.0507	0.0508	0.0595	0.0586	0.0576	0.0575	0.0573	0.0575
2000	Bimodal	Off by 2.5	80	0.0448	0.0446	0.0445	0.0448	0.0448	0.0446	0.0535	0.0514	0.0513	0.0512	0.0510	0.0508
2000	Chisq(5)	On-target	10	-0.0572	-0.0569	-0.0573	-0.0569	-0.0569	-0.0573	0.0906	0.0886	0.0880	0.0879	0.0879	0.0881
2000	Chisq(5)	On-target	20	-0.0594	-0.0602	-0.0606	-0.0602	-0.0604	-0.0604	0.0821	0.0803	0.0805	0.0798	0.0799	0.0800
2000	Chisq(5)	On-target	40	-0.0531	-0.0525	-0.0523	-0.0522	-0.0525	-0.0524	0.0723	0.0708	0.0696	0.0699	0.0700	0.0698
2000	Chisq(5)	On-target	60	-0.0502	-0.0494	-0.0495	-0.0494	-0.0495	-0.0496	0.0690	0.0678	0.0676	0.0675	0.0674	0.0675
2000	Chisq(5)	On-target	80	-0.0428	-0.0432	-0.0428	-0.0430	-0.0431	-0.0432	0.0631	0.0630	0.0624	0.0623	0.0623	0.0622
2000	Chisq(5)	Off by 0.5	10	-0.0185	-0.0172	-0.0170	-0.0176	-0.0177	-0.0174	0.0741	0.0712	0.0705	0.0703	0.0699	0.0699
2000	Chisq(5)	Off by 0.5	20	-0.0239	-0.0226	-0.0238	-0.0234	-0.0231	-0.0234	0.0631	0.0591	0.0604	0.0598	0.0593	0.0595
2000	Chisq(5)	Off by 0.5	40	-0.0300	-0.0315	-0.0308	-0.0309	-0.0306	-0.0309	0.0599	0.0592	0.0584	0.0581	0.0582	0.0580
2000	Chisq(5)	Off by 0.5	60	-0.0336	-0.0331	-0.0330	-0.0330	-0.0328	-0.0330	0.0607	0.0601	0.0596	0.0593	0.0592	0.0593
2000	Chisq(5)	Off by 0.5	80	-0.0313	-0.0310	-0.0311	-0.0310	-0.0309	-0.0311	0.0567	0.0556	0.0554	0.0552	0.0551	0.0551
2000	Chisq(5)	Off by 1.0	10	0.0295	0.0292	0.0284	0.0288	0.0288	0.0284	0.0852	0.0809	0.0806	0.0797	0.0795	0.0795
2000	Chisq(5)	Off by 1.0	20	0.0160	0.0154	0.0157	0.0154	0.0158	0.0157	0.0646	0.0625	0.0611	0.0609	0.0607	0.0607
2000	Chisq(5)	Off by 1.0	40	-0.0056	-0.0054	-0.0050	-0.0051	-0.0049	-0.0051	0.0542	0.0527	0.0521	0.0518	0.0518	0.0515
2000	Chisq(5)	Off by 1.0	60	-0.0126	-0.0120	-0.0122	-0.0122	-0.0120	-0.0121	0.0528	0.0514	0.0512	0.0512	0.0510	0.0509
2000	Chisq(5)	Off by 1.0	80	-0.0172	-0.0169	-0.0168	-0.0169	-0.0168	-0.0170	0.0547	0.0535	0.0534	0.0533	0.0531	0.0531
2000	Chisq(5)	Off by 2.5	10	0.2419	0.2414	0.2407	0.2420	0.2413	0.2416	0.2637	0.2610	0.2601	0.2613	0.2603	0.2605
2000	Chisq(5)	Off by 2.5	20	0.1781	0.1805	0.1785	0.1786	0.1792	0.1792	0.1940	0.1950	0.1927	0.1928	0.1933	0.1934
2000	Chisq(5)	Off by 2.5	40	0.1140	0.1140	0.1140	0.1139	0.1144	0.1141	0.1315	0.1298	0.1300	0.1296	0.1301	0.1296
2000	Chisq(5)	Off by 2.5	60	0.0768	0.0767	0.0763	0.0768	0.0767	0.0769	0.0944	0.0940	0.0935	0.0938	0.0935	0.0937
2000	Chisq(5)	Off by 2.5	80	0.0564	0.0569	0.0569	0.0570	0.0571	0.0571	0.0785	0.0778	0.0775	0.0774	0.0775	0.0777

Table 6 Bias and RMSE for Variance by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	-0.0004	-0.0005	-0.0007	-0.0006	-0.0007	-0.0006	0.0323	0.0308	0.0300	0.0301	0.0300	0.0300
10000	Normal	On-target	20	-0.0007	-0.0010	-0.0010	-0.0008	-0.0008	-0.0009	0.0234	0.0216	0.0214	0.0209	0.0211	0.0209
10000	Normal	On-target	40	-0.0001	0.0000	0.0001	0.0002	0.0000	0.0001	0.0197	0.0186	0.0183	0.0181	0.0182	0.0181
10000	Normal	On-target	60	0.0004	0.0005	0.0006	0.0006	0.0005	0.0006	0.0184	0.0176	0.0170	0.0171	0.0169	0.0169
10000	Normal	On-target	80	0.0016	0.0016	0.0014	0.0017	0.0015	0.0016	0.0169	0.0168	0.0166	0.0163	0.0164	0.0164
10000	Normal	Off by 0.5	10	0.0002	-0.0001	0.0002	0.0003	0.0001	0.0002	0.0315	0.0294	0.0292	0.0289	0.0290	0.0289
10000	Normal	Off by 0.5	20	-0.0003	-0.0002	-0.0003	-0.0001	-0.0002	-0.0002	0.0231	0.0219	0.0211	0.0214	0.0211	0.0211
10000	Normal	Off by 0.5	40	-0.0006	-0.0008	-0.0007	-0.0006	-0.0006	-0.0006	0.0193	0.0183	0.0182	0.0179	0.0178	0.0178
10000	Normal	Off by 0.5	60	0.0006	0.0009	0.0008	0.0008	0.0008	0.0008	0.0176	0.0171	0.0167	0.0168	0.0166	0.0166
10000	Normal	Off by 0.5	80	0.0015	0.0016	0.0017	0.0019	0.0019	0.0018	0.0176	0.0170	0.0169	0.0169	0.0169	0.0167
10000	Normal	Off by 1.0	10	-0.0010	-0.0012	-0.0012	-0.0011	-0.0011	-0.0011	0.0323	0.0310	0.0311	0.0306	0.0306	0.0304
10000	Normal	Off by 1.0	20	-0.0001	0.0004	0.0004	0.0004	0.0003	0.0004	0.0235	0.0221	0.0219	0.0215	0.0214	0.0213
10000	Normal	Off by 1.0	40	-0.0005	-0.0003	-0.0003	-0.0003	-0.0003	-0.0004	0.0205	0.0195	0.0190	0.0188	0.0188	0.0188
10000	Normal	Off by 1.0	60	0.0008	0.0006	0.0004	0.0006	0.0006	0.0006	0.0181	0.0171	0.0169	0.0167	0.0167	0.0166
10000	Normal	Off by 1.0	80	0.0015	0.0015	0.0013	0.0014	0.0015	0.0014	0.0180	0.0172	0.0169	0.0169	0.0169	0.0169
10000	Normal	Off by 2.5	10	-0.0002	-0.0003	0.0002	-0.0001	-0.0003	-0.0002	0.0370	0.0360	0.0357	0.0352	0.0352	0.0353
10000	Normal	Off by 2.5	20	0.0002	-0.0001	-0.0005	-0.0005	-0.0004	-0.0005	0.0273	0.0257	0.0251	0.0248	0.0248	0.0248
10000	Normal	Off by 2.5	40	-0.0008	-0.0006	-0.0004	-0.0003	-0.0005	-0.0005	0.0218	0.0205	0.0203	0.0201	0.0203	0.0202
10000	Normal	Off by 2.5	60	-0.0003	-0.0004	-0.0005	-0.0003	-0.0003	-0.0005	0.0203	0.0190	0.0189	0.0185	0.0185	0.0184
10000	Normal	Off by 2.5	80	0.0007	0.0004	0.0005	0.0005	0.0004	0.0005	0.0190	0.0184	0.0177	0.0177	0.0177	0.0177
10000	Uniform	On-target	10	0.0277	0.0270	0.0273	0.0275	0.0273	0.0272	0.0408	0.0391	0.0391	0.0391	0.0389	0.0387
10000	Uniform	On-target	20	0.0235	0.0230	0.0227	0.0230	0.0228	0.0229	0.0319	0.0304	0.0300	0.0300	0.0298	0.0298
10000	Uniform	On-target	40	0.0158	0.0156	0.0157	0.0157	0.0159	0.0158	0.0229	0.0219	0.0217	0.0217	0.0217	0.0217
10000	Uniform	On-target	60	0.0138	0.0138	0.0138	0.0138	0.0139	0.0137	0.0203	0.0196	0.0194	0.0193	0.0193	0.0192
10000	Uniform	On-target	80	0.0139	0.0136	0.0135	0.0135	0.0136	0.0135	0.0195	0.0190	0.0188	0.0186	0.0186	0.0186
10000	Uniform	Off by 0.5	10	0.0302	0.0303	0.0304	0.0306	0.0306	0.0305	0.0430	0.0418	0.0412	0.0410	0.0412	0.0411
10000	Uniform	Off by 0.5	20	0.0255	0.0256	0.0254	0.0257	0.0256	0.0255	0.0333	0.0324	0.0320	0.0321	0.0320	0.0319
10000	Uniform	Off by 0.5	40	0.0178	0.0177	0.0176	0.0178	0.0178	0.0178	0.0248	0.0239	0.0235	0.0236	0.0235	0.0235
10000	Uniform	Off by 0.5	60	0.0146	0.0150	0.0151	0.0150	0.0151	0.0151	0.0209	0.0207	0.0205	0.0204	0.0204	0.0204
10000	Uniform	Off by 0.5	80	0.0149	0.0149	0.0150	0.0149	0.0150	0.0149	0.0203	0.0198	0.0198	0.0196	0.0196	0.0195
10000	Uniform	Off by 1.0	10	0.0345	0.0344	0.0344	0.0347	0.0345	0.0344	0.0458	0.0447	0.0444	0.0444	0.0443	0.0441
10000	Uniform	Off by 1.0	20	0.0296	0.0301	0.0304	0.0302	0.0302	0.0302	0.0367	0.0366	0.0364	0.0362	0.0362	0.0361
10000	Uniform	Off by 1.0	40	0.0224	0.0230	0.0227	0.0227	0.0229	0.0227	0.0282	0.0276	0.0274	0.0273	0.0275	0.0272
10000	Uniform	Off by 1.0	60	0.0190	0.0190	0.0192	0.0190	0.0191	0.0190	0.0244	0.0238	0.0237	0.0234	0.0235	0.0234
10000	Uniform	Off by 1.0	80	0.0182	0.0182	0.0182	0.0183	0.0183	0.0184	0.0231	0.0227	0.0226	0.0225	0.0226	0.0226
10000	Uniform	Off by 2.5	10	0.0245	0.0239	0.0242	0.0241	0.0241	0.0240	0.0439	0.0419	0.0414	0.0411	0.0412	0.0410
10000	Uniform	Off by 2.5	20	0.0377	0.0381	0.0380	0.0380	0.0380	0.0382	0.0463	0.0459	0.0454	0.0455	0.0454	0.0454
10000	Uniform	Off by 2.5	40	0.0451	0.0458	0.0454	0.0455	0.0454	0.0454	0.0497	0.0497	0.0494	0.0492	0.0491	0.0491
10000	Uniform	Off by 2.5	60	0.0452	0.0451	0.0453	0.0451	0.0451	0.0450	0.0486	0.0481	0.0481	0.0478	0.0478	0.0477
10000	Uniform	Off by 2.5	80	0.0429	0.0429	0.0431	0.0430	0.0430	0.0429	0.0463	0.0459	0.0460	0.0457	0.0458	0.0457
10000	Bimodal	On-target	10	0.0277	0.0280	0.0279	0.0280	0.0280	0.0278	0.0397	0.0385	0.0382	0.0381	0.0381	0.0378
10000	Bimodal	On-target	20	0.0244	0.0239	0.0237	0.0239	0.0239	0.0239	0.0307	0.0297	0.0292	0.0292	0.0291	0.0290

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	0.0155	0.0159	0.0160	0.0158	0.0158	0.0159	0.0216	0.0207	0.0207	0.0205	0.0205	0.0205
10000	Bimodal	On-target	60	0.0125	0.0125	0.0126	0.0125	0.0125	0.0125	0.0173	0.0167	0.0165	0.0164	0.0163	0.0163
10000	Bimodal	On-target	80	0.0111	0.0111	0.0110	0.0112	0.0111	0.0111	0.0150	0.0145	0.0142	0.0142	0.0142	0.0141
10000	Bimodal	Off by 0.5	10	0.0304	0.0297	0.0305	0.0300	0.0300	0.0300	0.0428	0.0404	0.0408	0.0403	0.0400	0.0399
10000	Bimodal	Off by 0.5	20	0.0263	0.0269	0.0268	0.0266	0.0265	0.0266	0.0329	0.0322	0.0317	0.0314	0.0313	0.0313
10000	Bimodal	Off by 0.5	40	0.0173	0.0177	0.0176	0.0175	0.0175	0.0175	0.0233	0.0224	0.0222	0.0220	0.0220	0.0220
10000	Bimodal	Off by 0.5	60	0.0136	0.0138	0.0136	0.0137	0.0137	0.0137	0.0182	0.0176	0.0174	0.0174	0.0173	0.0172
10000	Bimodal	Off by 0.5	80	0.0124	0.0126	0.0124	0.0124	0.0125	0.0124	0.0162	0.0156	0.0154	0.0153	0.0153	0.0153
10000	Bimodal	Off by 1.0	10	0.0355	0.0346	0.0349	0.0351	0.0351	0.0348	0.0459	0.0439	0.0440	0.0438	0.0439	0.0437
10000	Bimodal	Off by 1.0	20	0.0334	0.0328	0.0330	0.0330	0.0330	0.0330	0.0388	0.0374	0.0374	0.0372	0.0371	0.0371
10000	Bimodal	Off by 1.0	40	0.0233	0.0230	0.0233	0.0232	0.0230	0.0229	0.0278	0.0269	0.0270	0.0268	0.0266	0.0266
10000	Bimodal	Off by 1.0	60	0.0183	0.0183	0.0183	0.0183	0.0183	0.0182	0.0219	0.0214	0.0212	0.0211	0.0211	0.0209
10000	Bimodal	Off by 1.0	80	0.0158	0.0159	0.0157	0.0159	0.0158	0.0157	0.0191	0.0187	0.0185	0.0185	0.0184	0.0183
10000	Bimodal	Off by 2.5	10	0.0276	0.0271	0.0271	0.0272	0.0273	0.0275	0.0438	0.0421	0.0420	0.0416	0.0417	0.0417
10000	Bimodal	Off by 2.5	20	0.0454	0.0445	0.0445	0.0445	0.0446	0.0444	0.0516	0.0499	0.0498	0.0497	0.0497	0.0496
10000	Bimodal	Off by 2.5	40	0.0514	0.0516	0.0513	0.0513	0.0514	0.0512	0.0548	0.0542	0.0539	0.0538	0.0539	0.0536
10000	Bimodal	Off by 2.5	60	0.0482	0.0483	0.0483	0.0483	0.0484	0.0483	0.0506	0.0504	0.0503	0.0502	0.0503	0.0502
10000	Bimodal	Off by 2.5	80	0.0434	0.0433	0.0436	0.0433	0.0433	0.0433	0.0455	0.0450	0.0451	0.0448	0.0448	0.0447
10000	Chisq(5)	On-target	10	-0.0591	-0.0596	-0.0592	-0.0594	-0.0594	-0.0593	0.0677	0.0671	0.0668	0.0669	0.0668	0.0667
10000	Chisq(5)	On-target	20	-0.0562	-0.0565	-0.0566	-0.0567	-0.0566	-0.0566	0.0624	0.0624	0.0624	0.0624	0.0623	0.0623
10000	Chisq(5)	On-target	40	-0.0526	-0.0523	-0.0526	-0.0526	-0.0524	-0.0526	0.0592	0.0586	0.0587	0.0587	0.0586	0.0587
10000	Chisq(5)	On-target	60	-0.0497	-0.0495	-0.0492	-0.0493	-0.0494	-0.0494	0.0564	0.0561	0.0558	0.0558	0.0559	0.0558
10000	Chisq(5)	On-target	80	-0.0460	-0.0458	-0.0458	-0.0458	-0.0458	-0.0458	0.0533	0.0529	0.0527	0.0528	0.0528	0.0528
10000	Chisq(5)	Off by 0.5	10	-0.0154	-0.0165	-0.0160	-0.0163	-0.0162	-0.0159	0.0372	0.0369	0.0368	0.0361	0.0361	0.0360
10000	Chisq(5)	Off by 0.5	20	-0.0241	-0.0242	-0.0243	-0.0242	-0.0243	-0.0242	0.0375	0.0368	0.0368	0.0365	0.0366	0.0365
10000	Chisq(5)	Off by 0.5	40	-0.0299	-0.0300	-0.0301	-0.0301	-0.0299	-0.0299	0.0408	0.0406	0.0405	0.0404	0.0403	0.0402
10000	Chisq(5)	Off by 0.5	60	-0.0326	-0.0325	-0.0324	-0.0324	-0.0324	-0.0325	0.0426	0.0421	0.0420	0.0420	0.0419	0.0419
10000	Chisq(5)	Off by 0.5	80	-0.0327	-0.0327	-0.0326	-0.0326	-0.0327	-0.0327	0.0423	0.0421	0.0419	0.0420	0.0420	0.0420
10000	Chisq(5)	Off by 1.0	10	0.0299	0.0296	0.0302	0.0300	0.0300	0.0298	0.0468	0.0452	0.0456	0.0453	0.0452	0.0451
10000	Chisq(5)	Off by 1.0	20	0.0121	0.0122	0.0122	0.0122	0.0122	0.0122	0.0327	0.0316	0.0316	0.0313	0.0313	0.0312
10000	Chisq(5)	Off by 1.0	40	-0.0064	-0.0063	-0.0063	-0.0063	-0.0062	-0.0062	0.0286	0.0281	0.0280	0.0279	0.0279	0.0279
10000	Chisq(5)	Off by 1.0	60	-0.0152	-0.0151	-0.0150	-0.0151	-0.0150	-0.0151	0.0314	0.0313	0.0309	0.0309	0.0310	0.0309
10000	Chisq(5)	Off by 1.0	80	-0.0189	-0.0192	-0.0191	-0.0190	-0.0190	-0.0191	0.0328	0.0325	0.0324	0.0324	0.0324	0.0324
10000	Chisq(5)	Off by 2.5	10	0.2423	0.2422	0.2428	0.2425	0.2425	0.2422	0.2468	0.2466	0.2470	0.2467	0.2467	0.2464
10000	Chisq(5)	Off by 2.5	20	0.1805	0.1806	0.1805	0.1809	0.1808	0.1809	0.1843	0.1842	0.1840	0.1843	0.1842	0.1843
10000	Chisq(5)	Off by 2.5	40	0.1129	0.1129	0.1126	0.1128	0.1124	0.1126	0.1170	0.1168	0.1165	0.1166	0.1163	0.1165
10000	Chisq(5)	Off by 2.5	60	0.0757	0.0762	0.0761	0.0759	0.0760	0.0761	0.0814	0.0817	0.0816	0.0814	0.0814	0.0815
10000	Chisq(5)	Off by 2.5	80	0.0553	0.0553	0.0553	0.0553	0.0552	0.0553	0.0623	0.0618	0.0619	0.0618	0.0616	0.0617

Table 7 Bias and RMSE for Percentage in the Top Band by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0018	0.0025	0.0013	0.0018	0.0017	0.0017	0.0435	0.0405	0.0387	0.0388	0.0387	0.0381
200	Normal	On-target	20	0.0036	0.0029	0.0028	0.0028	0.0030	0.0032	0.0377	0.0338	0.0323	0.0317	0.0322	0.0318
200	Normal	On-target	40	0.0033	0.0025	0.0020	0.0026	0.0024	0.0029	0.0331	0.0296	0.0291	0.0283	0.0282	0.0282
200	Normal	On-target	60	0.0021	0.0025	0.0027	0.0028	0.0029	0.0026	0.0314	0.0287	0.0282	0.0277	0.0277	0.0274
200	Normal	On-target	80	0.0018	0.0025	0.0024	0.0024	0.0026	0.0026	0.0307	0.0278	0.0276	0.0270	0.0268	0.0268
200	Normal	Off by 0.5	10	0.0001	-0.0004	-0.0011	-0.0008	-0.0007	-0.0006	0.0432	0.0390	0.0379	0.0377	0.0374	0.0372
200	Normal	Off by 0.5	20	0.0001	0.0011	0.0012	0.0005	0.0003	0.0006	0.0362	0.0322	0.0316	0.0309	0.0307	0.0309
200	Normal	Off by 0.5	40	0.0012	0.0008	0.0008	0.0007	0.0006	0.0008	0.0323	0.0290	0.0286	0.0280	0.0274	0.0277
200	Normal	Off by 0.5	60	0.0008	0.0014	0.0009	0.0004	0.0006	0.0007	0.0316	0.0285	0.0282	0.0274	0.0272	0.0271
200	Normal	Off by 0.5	80	0.0006	0.0007	0.0005	0.0007	0.0007	0.0007	0.0293	0.0277	0.0274	0.0267	0.0265	0.0264
200	Normal	Off by 1.0	10	-0.0030	-0.0036	-0.0035	-0.0031	-0.0030	-0.0032	0.0447	0.0401	0.0402	0.0401	0.0395	0.0395
200	Normal	Off by 1.0	20	-0.0006	-0.0013	-0.0008	-0.0012	-0.0014	-0.0012	0.0364	0.0330	0.0324	0.0314	0.0313	0.0308
200	Normal	Off by 1.0	40	-0.0026	-0.0013	-0.0016	-0.0019	-0.0016	-0.0014	0.0331	0.0298	0.0287	0.0282	0.0280	0.0279
200	Normal	Off by 1.0	60	-0.0016	-0.0015	-0.0017	-0.0019	-0.0017	-0.0021	0.0320	0.0286	0.0285	0.0278	0.0281	0.0278
200	Normal	Off by 1.0	80	-0.0020	-0.0016	-0.0019	-0.0021	-0.0021	-0.0021	0.0306	0.0271	0.0272	0.0264	0.0264	0.0263
200	Uniform	On-target	10	-0.0416	-0.0417	-0.0419	-0.0421	-0.0419	-0.0420	0.0592	0.0568	0.0558	0.0557	0.0553	0.0556
200	Uniform	On-target	20	-0.0349	-0.0360	-0.0352	-0.0355	-0.0353	-0.0354	0.0482	0.0466	0.0459	0.0457	0.0452	0.0453
200	Uniform	On-target	40	-0.0289	-0.0289	-0.0286	-0.0283	-0.0286	-0.0286	0.0425	0.0397	0.0393	0.0387	0.0387	0.0385
200	Uniform	On-target	60	-0.0211	-0.0217	-0.0215	-0.0220	-0.0217	-0.0216	0.0372	0.0353	0.0342	0.0345	0.0338	0.0338
200	Uniform	On-target	80	-0.0170	-0.0174	-0.0170	-0.0173	-0.0172	-0.0172	0.0359	0.0335	0.0329	0.0326	0.0324	0.0322
200	Uniform	Off by 0.5	10	-0.0443	-0.0438	-0.0445	-0.0442	-0.0441	-0.0441	0.0590	0.0567	0.0570	0.0564	0.0561	0.0561
200	Uniform	Off by 0.5	20	-0.0371	-0.0368	-0.0368	-0.0373	-0.0376	-0.0375	0.0498	0.0467	0.0465	0.0462	0.0461	0.0462
200	Uniform	Off by 0.5	40	-0.0297	-0.0302	-0.0301	-0.0302	-0.0302	-0.0299	0.0427	0.0406	0.0399	0.0398	0.0397	0.0397
200	Uniform	Off by 0.5	60	-0.0234	-0.0238	-0.0235	-0.0234	-0.0233	-0.0235	0.0381	0.0359	0.0353	0.0351	0.0346	0.0347
200	Uniform	Off by 0.5	80	-0.0183	-0.0186	-0.0183	-0.0186	-0.0187	-0.0188	0.0361	0.0338	0.0333	0.0332	0.0329	0.0328
200	Uniform	Off by 1.0	10	-0.0458	-0.0449	-0.0456	-0.0454	-0.0456	-0.0457	0.0612	0.0576	0.0573	0.0570	0.0569	0.0570
200	Uniform	Off by 1.0	20	-0.0392	-0.0406	-0.0399	-0.0399	-0.0397	-0.0399	0.0510	0.0499	0.0491	0.0486	0.0485	0.0484
200	Uniform	Off by 1.0	40	-0.0326	-0.0327	-0.0326	-0.0325	-0.0327	-0.0324	0.0446	0.0428	0.0422	0.0417	0.0416	0.0414
200	Uniform	Off by 1.0	60	-0.0249	-0.0254	-0.0249	-0.0252	-0.0253	-0.0252	0.0388	0.0370	0.0360	0.0359	0.0359	0.0358
200	Uniform	Off by 1.0	80	-0.0223	-0.0218	-0.0213	-0.0217	-0.0215	-0.0214	0.0374	0.0349	0.0341	0.0344	0.0339	0.0339
200	Bimodal	On-target	10	-0.0202	-0.0200	-0.0202	-0.0200	-0.0199	-0.0203	0.0458	0.0419	0.0413	0.0405	0.0406	0.0404
200	Bimodal	On-target	20	-0.0097	-0.0095	-0.0093	-0.0101	-0.0096	-0.0098	0.0373	0.0328	0.0315	0.0316	0.0311	0.0312
200	Bimodal	On-target	40	-0.0006	0.0005	0.0002	0.0004	0.0004	0.0004	0.0322	0.0288	0.0283	0.0279	0.0273	0.0273
200	Bimodal	On-target	60	0.0081	0.0086	0.0087	0.0085	0.0086	0.0086	0.0332	0.0300	0.0287	0.0285	0.0285	0.0283
200	Bimodal	On-target	80	0.0120	0.0129	0.0125	0.0129	0.0128	0.0127	0.0343	0.0312	0.0302	0.0298	0.0296	0.0293
200	Bimodal	Off by 0.5	10	-0.0219	-0.0228	-0.0233	-0.0229	-0.0230	-0.0230	0.0439	0.0421	0.0415	0.0405	0.0403	0.0404
200	Bimodal	Off by 0.5	20	-0.0119	-0.0118	-0.0126	-0.0125	-0.0126	-0.0126	0.0365	0.0327	0.0322	0.0317	0.0311	0.0313
200	Bimodal	Off by 0.5	40	-0.0021	-0.0021	-0.0015	-0.0019	-0.0021	-0.0018	0.0325	0.0283	0.0279	0.0274	0.0272	0.0272
200	Bimodal	Off by 0.5	60	0.0062	0.0062	0.0065	0.0062	0.0065	0.0062	0.0324	0.0293	0.0285	0.0283	0.0280	0.0280
200	Bimodal	Off by 0.5	80	0.0116	0.0112	0.0106	0.0110	0.0111	0.0110	0.0335	0.0305	0.0293	0.0289	0.0289	0.0288
200	Bimodal	Off by 1.0	10	-0.0239	-0.0252	-0.0250	-0.0253	-0.0251	-0.0253	0.0456	0.0427	0.0425	0.0419	0.0416	0.0417
200	Bimodal	Off by 1.0	20	-0.0146	-0.0161	-0.0156	-0.0159	-0.0157	-0.0155	0.0379	0.0355	0.0347	0.0337	0.0335	0.0334

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 1.0	40	-0.0044	-0.0043	-0.0048	-0.0048	-0.0045	-0.0048	0.0323	0.0282	0.0279	0.0280	0.0273	0.0273
200	Bimodal	Off by 1.0	60	0.0030	0.0035	0.0035	0.0035	0.0032	0.0032	0.0322	0.0287	0.0279	0.0274	0.0273	0.0269
200	Bimodal	Off by 1.0	80	0.0080	0.0080	0.0081	0.0083	0.0083	0.0082	0.0315	0.0295	0.0284	0.0277	0.0276	0.0276
200	Chisq(5)	On-target	10	0.0225	0.0233	0.0225	0.0228	0.0230	0.0230	0.0531	0.0513	0.0497	0.0493	0.0493	0.0491
200	Chisq(5)	On-target	20	0.0162	0.0163	0.0164	0.0165	0.0165	0.0166	0.0426	0.0388	0.0380	0.0378	0.0374	0.0373
200	Chisq(5)	On-target	40	0.0101	0.0107	0.0104	0.0105	0.0104	0.0106	0.0359	0.0333	0.0322	0.0322	0.0320	0.0319
200	Chisq(5)	On-target	60	0.0086	0.0076	0.0085	0.0081	0.0080	0.0081	0.0342	0.0316	0.0309	0.0307	0.0307	0.0305
200	Chisq(5)	On-target	80	0.0065	0.0069	0.0064	0.0068	0.0067	0.0065	0.0329	0.0313	0.0304	0.0305	0.0302	0.0302
200	Chisq(5)	Off by 0.5	10	0.0255	0.0253	0.0255	0.0251	0.0251	0.0251	0.0531	0.0508	0.0504	0.0498	0.0494	0.0492
200	Chisq(5)	Off by 0.5	20	0.0173	0.0171	0.0165	0.0171	0.0170	0.0173	0.0418	0.0387	0.0376	0.0376	0.0370	0.0371
200	Chisq(5)	Off by 0.5	40	0.0096	0.0100	0.0100	0.0099	0.0096	0.0097	0.0351	0.0328	0.0322	0.0314	0.0314	0.0313
200	Chisq(5)	Off by 0.5	60	0.0074	0.0071	0.0072	0.0070	0.0071	0.0071	0.0324	0.0310	0.0310	0.0304	0.0300	0.0299
200	Chisq(5)	Off by 0.5	80	0.0048	0.0056	0.0055	0.0053	0.0053	0.0054	0.0331	0.0301	0.0303	0.0301	0.0297	0.0298
200	Chisq(5)	Off by 1.0	10	0.0280	0.0263	0.0260	0.0264	0.0263	0.0264	0.0560	0.0509	0.0500	0.0491	0.0490	0.0493
200	Chisq(5)	Off by 1.0	20	0.0172	0.0160	0.0169	0.0169	0.0170	0.0171	0.0418	0.0386	0.0383	0.0378	0.0376	0.0374
200	Chisq(5)	Off by 1.0	40	0.0083	0.0084	0.0084	0.0083	0.0083	0.0084	0.0341	0.0327	0.0320	0.0316	0.0316	0.0313
200	Chisq(5)	Off by 1.0	60	0.0061	0.0056	0.0060	0.0056	0.0058	0.0059	0.0328	0.0305	0.0304	0.0300	0.0298	0.0298
200	Chisq(5)	Off by 1.0	80	0.0034	0.0037	0.0036	0.0038	0.0038	0.0036	0.0324	0.0298	0.0300	0.0295	0.0297	0.0296

Table 8 Bias and RMSE for Percentage in the Top Band by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0009	0.0012	0.0012	0.0012	0.0009	0.0011	0.0142	0.0132	0.0130	0.0129	0.0127	0.0128
2000	Normal	On-target	20	0.0003	0.0004	0.0006	0.0006	0.0006	0.0006	0.0120	0.0110	0.0108	0.0107	0.0106	0.0106
2000	Normal	On-target	40	0.0004	0.0003	0.0003	0.0004	0.0003	0.0004	0.0111	0.0102	0.0100	0.0098	0.0097	0.0098
2000	Normal	On-target	60	0.0004	0.0004	0.0004	0.0005	0.0005	0.0004	0.0104	0.0098	0.0096	0.0095	0.0094	0.0094
2000	Normal	On-target	80	0.0009	0.0009	0.0009	0.0008	0.0009	0.0009	0.0102	0.0095	0.0092	0.0091	0.0092	0.0090
2000	Normal	Off by 0.5	10	0.0012	0.0009	0.0007	0.0007	0.0006	0.0006	0.0137	0.0126	0.0123	0.0121	0.0119	0.0120
2000	Normal	Off by 0.5	20	0.0005	0.0001	0.0001	-0.0001	0.0001	0.0000	0.0130	0.0114	0.0111	0.0111	0.0110	0.0110
2000	Normal	Off by 0.5	40	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0112	0.0102	0.0100	0.0099	0.0098	0.0098
2000	Normal	Off by 0.5	60	0.0006	0.0003	0.0003	0.0003	0.0003	0.0002	0.0108	0.0099	0.0098	0.0097	0.0096	0.0095
2000	Normal	Off by 0.5	80	0.0012	0.0010	0.0009	0.0010	0.0010	0.0010	0.0110	0.0101	0.0099	0.0098	0.0100	0.0098
2000	Normal	Off by 1.0	10	0.0002	0.0001	0.0001	-0.0001	0.0000	0.0001	0.0147	0.0131	0.0132	0.0130	0.0129	0.0128
2000	Normal	Off by 1.0	20	-0.0004	0.0000	0.0001	0.0000	0.0000	0.0000	0.0124	0.0115	0.0113	0.0110	0.0109	0.0109
2000	Normal	Off by 1.0	40	-0.0002	-0.0004	-0.0001	-0.0002	-0.0003	-0.0003	0.0111	0.0102	0.0099	0.0098	0.0098	0.0098
2000	Normal	Off by 1.0	60	0.0001	0.0004	0.0003	0.0003	0.0003	0.0003	0.0109	0.0096	0.0095	0.0094	0.0094	0.0093
2000	Normal	Off by 1.0	80	-0.0003	-0.0005	-0.0003	-0.0004	-0.0004	-0.0004	0.0105	0.0097	0.0095	0.0094	0.0094	0.0093
2000	Normal	Off by 2.5	10	-0.0017	-0.0015	-0.0016	-0.0015	-0.0016	-0.0016	0.0166	0.0156	0.0154	0.0155	0.0155	0.0154
2000	Normal	Off by 2.5	20	-0.0016	-0.0009	-0.0013	-0.0012	-0.0013	-0.0013	0.0141	0.0128	0.0127	0.0124	0.0124	0.0123
2000	Normal	Off by 2.5	40	-0.0021	-0.0019	-0.0019	-0.0020	-0.0020	-0.0020	0.0119	0.0112	0.0109	0.0107	0.0107	0.0107
2000	Normal	Off by 2.5	60	-0.0012	-0.0013	-0.0012	-0.0014	-0.0013	-0.0014	0.0114	0.0103	0.0103	0.0102	0.0101	0.0102
2000	Normal	Off by 2.5	80	-0.0020	-0.0020	-0.0020	-0.0019	-0.0019	-0.0020	0.0108	0.0101	0.0099	0.0097	0.0097	0.0098
2000	Uniform	On-target	10	-0.0448	-0.0453	-0.0452	-0.0451	-0.0451	-0.0451	0.0467	0.0469	0.0468	0.0467	0.0467	0.0467
2000	Uniform	On-target	20	-0.0392	-0.0394	-0.0393	-0.0393	-0.0395	-0.0394	0.0409	0.0408	0.0407	0.0406	0.0408	0.0407
2000	Uniform	On-target	40	-0.0312	-0.0313	-0.0311	-0.0312	-0.0311	-0.0312	0.0330	0.0328	0.0326	0.0326	0.0326	0.0326
2000	Uniform	On-target	60	-0.0249	-0.0251	-0.0250	-0.0250	-0.0251	-0.0250	0.0270	0.0269	0.0268	0.0267	0.0268	0.0267
2000	Uniform	On-target	80	-0.0206	-0.0204	-0.0205	-0.0204	-0.0205	-0.0205	0.0230	0.0225	0.0224	0.0224	0.0224	0.0224
2000	Uniform	Off by 0.5	10	-0.0450	-0.0450	-0.0448	-0.0449	-0.0450	-0.0451	0.0469	0.0466	0.0464	0.0464	0.0464	0.0465
2000	Uniform	Off by 0.5	20	-0.0400	-0.0401	-0.0400	-0.0401	-0.0401	-0.0401	0.0416	0.0414	0.0412	0.0413	0.0413	0.0413
2000	Uniform	Off by 0.5	40	-0.0311	-0.0311	-0.0311	-0.0312	-0.0310	-0.0311	0.0328	0.0326	0.0326	0.0325	0.0323	0.0324
2000	Uniform	Off by 0.5	60	-0.0248	-0.0249	-0.0249	-0.0247	-0.0247	-0.0247	0.0269	0.0265	0.0266	0.0263	0.0263	0.0263
2000	Uniform	Off by 0.5	80	-0.0213	-0.0212	-0.0213	-0.0212	-0.0212	-0.0213	0.0238	0.0232	0.0233	0.0231	0.0232	0.0232
2000	Uniform	Off by 1.0	10	-0.0455	-0.0453	-0.0454	-0.0454	-0.0454	-0.0455	0.0472	0.0468	0.0468	0.0468	0.0468	0.0468
2000	Uniform	Off by 1.0	20	-0.0405	-0.0398	-0.0401	-0.0399	-0.0400	-0.0400	0.0422	0.0411	0.0413	0.0411	0.0412	0.0412
2000	Uniform	Off by 1.0	40	-0.0309	-0.0310	-0.0310	-0.0310	-0.0309	-0.0310	0.0326	0.0326	0.0324	0.0325	0.0324	0.0324
2000	Uniform	Off by 1.0	60	-0.0251	-0.0249	-0.0251	-0.0251	-0.0251	-0.0250	0.0272	0.0266	0.0267	0.0267	0.0267	0.0266
2000	Uniform	Off by 1.0	80	-0.0207	-0.0207	-0.0208	-0.0208	-0.0208	-0.0208	0.0231	0.0228	0.0227	0.0226	0.0226	0.0226
2000	Uniform	Off by 2.5	10	-0.0503	-0.0504	-0.0505	-0.0503	-0.0503	-0.0504	0.0524	0.0522	0.0522	0.0520	0.0520	0.0521
2000	Uniform	Off by 2.5	20	-0.0450	-0.0448	-0.0447	-0.0447	-0.0449	-0.0448	0.0467	0.0463	0.0461	0.0461	0.0463	0.0462
2000	Uniform	Off by 2.5	40	-0.0346	-0.0351	-0.0350	-0.0349	-0.0350	-0.0350	0.0364	0.0365	0.0365	0.0363	0.0363	0.0363
2000	Uniform	Off by 2.5	60	-0.0288	-0.0289	-0.0287	-0.0288	-0.0287	-0.0288	0.0306	0.0305	0.0302	0.0303	0.0303	0.0303
2000	Uniform	Off by 2.5	80	-0.0252	-0.0251	-0.0251	-0.0250	-0.0251	-0.0251	0.0270	0.0268	0.0267	0.0266	0.0267	0.0267
2000	Bimodal	On-target	10	-0.0196	-0.0200	-0.0199	-0.0199	-0.0200	-0.0199	0.0240	0.0238	0.0234	0.0233	0.0234	0.0233
2000	Bimodal	On-target	20	-0.0120	-0.0123	-0.0122	-0.0122	-0.0123	-0.0122	0.0168	0.0164	0.0161	0.0161	0.0159	0.0159

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	-0.0008	-0.0009	-0.0009	-0.0009	-0.0009	-0.0009	0.0118	0.0106	0.0104	0.0102	0.0101	0.0100
2000	Bimodal	On-target	60	0.0053	0.0054	0.0052	0.0054	0.0055	0.0054	0.0125	0.0114	0.0113	0.0112	0.0111	0.0111
2000	Bimodal	On-target	80	0.0103	0.0104	0.0103	0.0104	0.0104	0.0104	0.0146	0.0143	0.0141	0.0140	0.0140	0.0140
2000	Bimodal	Off by 0.5	10	-0.0207	-0.0205	-0.0203	-0.0204	-0.0205	-0.0206	0.0245	0.0237	0.0235	0.0234	0.0235	0.0235
2000	Bimodal	Off by 0.5	20	-0.0123	-0.0125	-0.0125	-0.0124	-0.0125	-0.0125	0.0170	0.0162	0.0160	0.0159	0.0158	0.0158
2000	Bimodal	Off by 0.5	40	-0.0010	-0.0012	-0.0013	-0.0012	-0.0011	-0.0012	0.0106	0.0096	0.0093	0.0092	0.0092	0.0092
2000	Bimodal	Off by 0.5	60	0.0062	0.0060	0.0058	0.0059	0.0058	0.0058	0.0122	0.0111	0.0106	0.0106	0.0105	0.0105
2000	Bimodal	Off by 0.5	80	0.0104	0.0105	0.0103	0.0104	0.0104	0.0104	0.0149	0.0140	0.0139	0.0138	0.0137	0.0138
2000	Bimodal	Off by 1.0	10	-0.0213	-0.0209	-0.0211	-0.0211	-0.0211	-0.0211	0.0251	0.0241	0.0241	0.0240	0.0240	0.0239
2000	Bimodal	Off by 1.0	20	-0.0139	-0.0136	-0.0136	-0.0136	-0.0136	-0.0136	0.0184	0.0173	0.0172	0.0171	0.0171	0.0171
2000	Bimodal	Off by 1.0	40	-0.0021	-0.0023	-0.0021	-0.0022	-0.0021	-0.0022	0.0115	0.0103	0.0100	0.0099	0.0097	0.0097
2000	Bimodal	Off by 1.0	60	0.0054	0.0050	0.0051	0.0052	0.0051	0.0052	0.0124	0.0113	0.0111	0.0112	0.0109	0.0110
2000	Bimodal	Off by 1.0	80	0.0099	0.0096	0.0097	0.0098	0.0097	0.0097	0.0149	0.0140	0.0140	0.0139	0.0138	0.0138
2000	Bimodal	Off by 2.5	10	-0.0255	-0.0254	-0.0253	-0.0254	-0.0256	-0.0255	0.0302	0.0296	0.0295	0.0294	0.0296	0.0295
2000	Bimodal	Off by 2.5	20	-0.0191	-0.0186	-0.0188	-0.0187	-0.0188	-0.0186	0.0231	0.0222	0.0222	0.0221	0.0221	0.0219
2000	Bimodal	Off by 2.5	40	-0.0068	-0.0065	-0.0067	-0.0066	-0.0067	-0.0066	0.0140	0.0129	0.0130	0.0128	0.0128	0.0127
2000	Bimodal	Off by 2.5	60	-0.0004	0.0000	-0.0002	-0.0001	-0.0001	-0.0002	0.0113	0.0104	0.0102	0.0102	0.0101	0.0100
2000	Bimodal	Off by 2.5	80	0.0054	0.0053	0.0051	0.0052	0.0052	0.0052	0.0125	0.0120	0.0118	0.0115	0.0116	0.0115
2000	Chisq(5)	On-target	10	0.0232	0.0234	0.0234	0.0236	0.0235	0.0235	0.0282	0.0274	0.0273	0.0273	0.0272	0.0272
2000	Chisq(5)	On-target	20	0.0149	0.0147	0.0145	0.0147	0.0147	0.0147	0.0198	0.0188	0.0185	0.0185	0.0185	0.0185
2000	Chisq(5)	On-target	40	0.0087	0.0086	0.0084	0.0086	0.0085	0.0085	0.0141	0.0134	0.0130	0.0131	0.0129	0.0130
2000	Chisq(5)	On-target	60	0.0061	0.0062	0.0061	0.0062	0.0062	0.0062	0.0119	0.0114	0.0112	0.0111	0.0111	0.0111
2000	Chisq(5)	On-target	80	0.0051	0.0048	0.0049	0.0050	0.0050	0.0049	0.0111	0.0104	0.0105	0.0103	0.0103	0.0103
2000	Chisq(5)	Off by 0.5	10	0.0266	0.0270	0.0270	0.0269	0.0269	0.0269	0.0307	0.0306	0.0305	0.0303	0.0302	0.0302
2000	Chisq(5)	Off by 0.5	20	0.0164	0.0167	0.0165	0.0166	0.0167	0.0167	0.0207	0.0200	0.0199	0.0199	0.0199	0.0199
2000	Chisq(5)	Off by 0.5	40	0.0091	0.0091	0.0092	0.0091	0.0092	0.0092	0.0141	0.0136	0.0134	0.0132	0.0133	0.0132
2000	Chisq(5)	Off by 0.5	60	0.0059	0.0059	0.0060	0.0060	0.0060	0.0060	0.0123	0.0117	0.0117	0.0116	0.0116	0.0115
2000	Chisq(5)	Off by 0.5	80	0.0048	0.0050	0.0048	0.0048	0.0048	0.0048	0.0109	0.0103	0.0104	0.0102	0.0102	0.0101
2000	Chisq(5)	Off by 1.0	10	0.0307	0.0314	0.0313	0.0312	0.0312	0.0312	0.0341	0.0340	0.0339	0.0336	0.0336	0.0336
2000	Chisq(5)	Off by 1.0	20	0.0195	0.0196	0.0196	0.0197	0.0197	0.0197	0.0231	0.0226	0.0224	0.0224	0.0224	0.0224
2000	Chisq(5)	Off by 1.0	40	0.0110	0.0108	0.0108	0.0108	0.0109	0.0107	0.0155	0.0147	0.0145	0.0145	0.0145	0.0144
2000	Chisq(5)	Off by 1.0	60	0.0074	0.0074	0.0073	0.0073	0.0074	0.0074	0.0128	0.0125	0.0122	0.0120	0.0121	0.0121
2000	Chisq(5)	Off by 1.0	80	0.0053	0.0052	0.0052	0.0052	0.0053	0.0051	0.0115	0.0109	0.0107	0.0107	0.0107	0.0106
2000	Chisq(5)	Off by 2.5	10	0.0422	0.0424	0.0421	0.0423	0.0423	0.0422	0.0454	0.0452	0.0448	0.0449	0.0449	0.0449
2000	Chisq(5)	Off by 2.5	20	0.0293	0.0292	0.0291	0.0290	0.0291	0.0292	0.0324	0.0319	0.0317	0.0316	0.0316	0.0317
2000	Chisq(5)	Off by 2.5	40	0.0158	0.0161	0.0158	0.0159	0.0159	0.0159	0.0198	0.0193	0.0192	0.0191	0.0190	0.0191
2000	Chisq(5)	Off by 2.5	60	0.0104	0.0106	0.0106	0.0105	0.0105	0.0105	0.0148	0.0144	0.0143	0.0141	0.0141	0.0141
2000	Chisq(5)	Off by 2.5	80	0.0073	0.0074	0.0073	0.0073	0.0074	0.0074	0.0127	0.0122	0.0121	0.0120	0.0120	0.0121

Table 9 Bias and RMSE for Percentage in the Top Band by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	0.0001	0.0001	0.0001	0.0000	0.0000	0.0001	0.0078	0.0075	0.0074	0.0073	0.0073	0.0073
10000	Normal	On-target	20	0.0001	0.0000	0.0000	0.0001	0.0001	0.0001	0.0070	0.0066	0.0065	0.0064	0.0064	0.0064
10000	Normal	On-target	40	0.0003	0.0004	0.0003	0.0003	0.0003	0.0003	0.0063	0.0060	0.0058	0.0058	0.0058	0.0058
10000	Normal	On-target	60	0.0003	0.0004	0.0003	0.0004	0.0003	0.0004	0.0059	0.0057	0.0056	0.0056	0.0055	0.0056
10000	Normal	On-target	80	0.0004	0.0004	0.0003	0.0003	0.0003	0.0004	0.0056	0.0052	0.0051	0.0051	0.0051	0.0051
10000	Normal	Off by 0.5	10	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0077	0.0071	0.0070	0.0070	0.0069	0.0069
10000	Normal	Off by 0.5	20	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0067	0.0064	0.0064	0.0063	0.0063	0.0064
10000	Normal	Off by 0.5	40	0.0002	0.0002	0.0001	0.0002	0.0002	0.0002	0.0063	0.0059	0.0058	0.0058	0.0057	0.0057
10000	Normal	Off by 0.5	60	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0058	0.0056	0.0055	0.0054	0.0054	0.0054
10000	Normal	Off by 0.5	80	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0057	0.0054	0.0053	0.0052	0.0053	0.0052
10000	Normal	Off by 1.0	10	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	0.0078	0.0073	0.0072	0.0072	0.0072	0.0071
10000	Normal	Off by 1.0	20	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001	0.0069	0.0065	0.0064	0.0064	0.0063	0.0063
10000	Normal	Off by 1.0	40	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001	0.0065	0.0062	0.0061	0.0060	0.0059	0.0059
10000	Normal	Off by 1.0	60	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0057	0.0056	0.0055	0.0054	0.0054	0.0054
10000	Normal	Off by 1.0	80	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0057	0.0054	0.0053	0.0053	0.0053	0.0053
10000	Normal	Off by 2.5	10	-0.0005	-0.0007	-0.0006	-0.0007	-0.0007	-0.0007	0.0089	0.0085	0.0084	0.0083	0.0083	0.0083
10000	Normal	Off by 2.5	20	-0.0003	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002	0.0074	0.0071	0.0069	0.0069	0.0069	0.0069
10000	Normal	Off by 2.5	40	-0.0003	-0.0003	-0.0004	-0.0003	-0.0003	-0.0003	0.0065	0.0062	0.0061	0.0060	0.0060	0.0060
10000	Normal	Off by 2.5	60	-0.0002	-0.0002	-0.0002	-0.0002	-0.0003	-0.0002	0.0060	0.0056	0.0056	0.0056	0.0055	0.0056
10000	Normal	Off by 2.5	80	-0.0003	-0.0005	-0.0003	-0.0004	-0.0004	-0.0004	0.0059	0.0055	0.0055	0.0055	0.0054	0.0054
10000	Uniform	On-target	10	-0.0449	-0.0449	-0.0449	-0.0449	-0.0449	-0.0450	0.0455	0.0455	0.0454	0.0454	0.0454	0.0454
10000	Uniform	On-target	20	-0.0399	-0.0399	-0.0399	-0.0399	-0.0399	-0.0399	0.0404	0.0403	0.0404	0.0403	0.0404	0.0403
10000	Uniform	On-target	40	-0.0315	-0.0316	-0.0315	-0.0315	-0.0315	-0.0315	0.0321	0.0321	0.0321	0.0320	0.0320	0.0320
10000	Uniform	On-target	60	-0.0255	-0.0256	-0.0256	-0.0256	-0.0255	-0.0256	0.0263	0.0262	0.0262	0.0262	0.0261	0.0262
10000	Uniform	On-target	80	-0.0211	-0.0211	-0.0211	-0.0211	-0.0211	-0.0211	0.0219	0.0218	0.0219	0.0218	0.0218	0.0218
10000	Uniform	Off by 0.5	10	-0.0456	-0.0456	-0.0455	-0.0455	-0.0455	-0.0455	0.0461	0.0461	0.0459	0.0460	0.0459	0.0459
10000	Uniform	Off by 0.5	20	-0.0400	-0.0401	-0.0400	-0.0400	-0.0401	-0.0400	0.0405	0.0405	0.0405	0.0405	0.0405	0.0404
10000	Uniform	Off by 0.5	40	-0.0314	-0.0314	-0.0315	-0.0314	-0.0315	-0.0314	0.0320	0.0319	0.0320	0.0319	0.0320	0.0319
10000	Uniform	Off by 0.5	60	-0.0254	-0.0254	-0.0254	-0.0254	-0.0254	-0.0254	0.0261	0.0260	0.0261	0.0260	0.0260	0.0260
10000	Uniform	Off by 0.5	80	-0.0210	-0.0210	-0.0210	-0.0210	-0.0210	-0.0210	0.0218	0.0218	0.0217	0.0217	0.0217	0.0217
10000	Uniform	Off by 1.0	10	-0.0459	-0.0459	-0.0460	-0.0460	-0.0460	-0.0460	0.0465	0.0464	0.0465	0.0464	0.0464	0.0465
10000	Uniform	Off by 1.0	20	-0.0405	-0.0404	-0.0404	-0.0403	-0.0404	-0.0404	0.0411	0.0408	0.0408	0.0408	0.0408	0.0408
10000	Uniform	Off by 1.0	40	-0.0316	-0.0315	-0.0315	-0.0316	-0.0316	-0.0316	0.0322	0.0321	0.0320	0.0321	0.0321	0.0321
10000	Uniform	Off by 1.0	60	-0.0256	-0.0255	-0.0255	-0.0255	-0.0256	-0.0256	0.0263	0.0262	0.0261	0.0261	0.0262	0.0262
10000	Uniform	Off by 1.0	80	-0.0213	-0.0214	-0.0214	-0.0213	-0.0213	-0.0213	0.0221	0.0221	0.0221	0.0220	0.0220	0.0220
10000	Uniform	Off by 2.5	10	-0.0491	-0.0491	-0.0491	-0.0492	-0.0491	-0.0491	0.0497	0.0497	0.0496	0.0497	0.0497	0.0497
10000	Uniform	Off by 2.5	20	-0.0430	-0.0431	-0.0431	-0.0431	-0.0431	-0.0430	0.0436	0.0435	0.0436	0.0435	0.0435	0.0435
10000	Uniform	Off by 2.5	40	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	0.0346	0.0345	0.0346	0.0346	0.0346	0.0346
10000	Uniform	Off by 2.5	60	-0.0279	-0.0279	-0.0279	-0.0280	-0.0279	-0.0279	0.0286	0.0285	0.0285	0.0286	0.0285	0.0285
10000	Uniform	Off by 2.5	80	-0.0236	-0.0237	-0.0237	-0.0236	-0.0237	-0.0236	0.0243	0.0244	0.0243	0.0243	0.0243	0.0243
10000	Bimodal	On-target	10	-0.0200	-0.0199	-0.0199	-0.0200	-0.0199	-0.0199	0.0213	0.0211	0.0211	0.0211	0.0211	0.0210
10000	Bimodal	On-target	20	-0.0125	-0.0126	-0.0126	-0.0126	-0.0126	-0.0126	0.0143	0.0142	0.0142	0.0142	0.0141	0.0141

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	-0.0014	-0.0014	-0.0013	-0.0014	-0.0014	-0.0014	0.0068	0.0063	0.0063	0.0062	0.0062	0.0062
10000	Bimodal	On-target	60	0.0054	0.0054	0.0056	0.0054	0.0055	0.0054	0.0086	0.0083	0.0084	0.0082	0.0083	0.0082
10000	Bimodal	On-target	80	0.0102	0.0102	0.0101	0.0102	0.0102	0.0102	0.0120	0.0118	0.0117	0.0118	0.0118	0.0117
10000	Bimodal	Off by 0.5	10	-0.0207	-0.0207	-0.0207	-0.0207	-0.0207	-0.0207	0.0220	0.0218	0.0218	0.0218	0.0218	0.0217
10000	Bimodal	Off by 0.5	20	-0.0129	-0.0128	-0.0129	-0.0128	-0.0129	-0.0129	0.0147	0.0144	0.0144	0.0143	0.0144	0.0143
10000	Bimodal	Off by 0.5	40	-0.0016	-0.0016	-0.0015	-0.0015	-0.0015	-0.0015	0.0067	0.0064	0.0062	0.0063	0.0062	0.0062
10000	Bimodal	Off by 0.5	60	0.0054	0.0055	0.0054	0.0055	0.0055	0.0054	0.0085	0.0083	0.0082	0.0082	0.0082	0.0081
10000	Bimodal	Off by 0.5	80	0.0100	0.0101	0.0100	0.0100	0.0100	0.0100	0.0119	0.0118	0.0117	0.0116	0.0116	0.0117
10000	Bimodal	Off by 1.0	10	-0.0212	-0.0213	-0.0212	-0.0212	-0.0211	-0.0212	0.0224	0.0223	0.0222	0.0223	0.0222	0.0222
10000	Bimodal	Off by 1.0	20	-0.0131	-0.0133	-0.0132	-0.0132	-0.0132	-0.0132	0.0147	0.0147	0.0147	0.0146	0.0146	0.0146
10000	Bimodal	Off by 1.0	40	-0.0018	-0.0018	-0.0019	-0.0019	-0.0018	-0.0019	0.0069	0.0064	0.0063	0.0064	0.0063	0.0063
10000	Bimodal	Off by 1.0	60	0.0051	0.0052	0.0051	0.0051	0.0052	0.0051	0.0082	0.0081	0.0080	0.0079	0.0080	0.0079
10000	Bimodal	Off by 1.0	80	0.0096	0.0097	0.0097	0.0097	0.0097	0.0096	0.0116	0.0114	0.0114	0.0114	0.0113	0.0113
10000	Bimodal	Off by 2.5	10	-0.0244	-0.0244	-0.0244	-0.0244	-0.0244	-0.0243	0.0257	0.0255	0.0255	0.0255	0.0255	0.0254
10000	Bimodal	Off by 2.5	20	-0.0169	-0.0171	-0.0170	-0.0170	-0.0170	-0.0171	0.0184	0.0184	0.0183	0.0183	0.0183	0.0183
10000	Bimodal	Off by 2.5	40	-0.0057	-0.0056	-0.0056	-0.0056	-0.0056	-0.0057	0.0088	0.0084	0.0084	0.0084	0.0084	0.0084
10000	Bimodal	Off by 2.5	60	0.0014	0.0016	0.0016	0.0017	0.0016	0.0016	0.0070	0.0067	0.0066	0.0065	0.0065	0.0065
10000	Bimodal	Off by 2.5	80	0.0064	0.0064	0.0065	0.0065	0.0064	0.0065	0.0093	0.0090	0.0090	0.0089	0.0089	0.0090
10000	Chisq(5)	On-target	10	0.0236	0.0235	0.0236	0.0236	0.0236	0.0236	0.0252	0.0249	0.0249	0.0249	0.0249	0.0249
10000	Chisq(5)	On-target	20	0.0156	0.0155	0.0155	0.0155	0.0154	0.0155	0.0170	0.0169	0.0168	0.0168	0.0168	0.0168
10000	Chisq(5)	On-target	40	0.0088	0.0087	0.0088	0.0088	0.0088	0.0088	0.0108	0.0106	0.0106	0.0105	0.0105	0.0105
10000	Chisq(5)	On-target	60	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0083	0.0081	0.0081	0.0080	0.0080	0.0080
10000	Chisq(5)	On-target	80	0.0048	0.0049	0.0049	0.0049	0.0049	0.0049	0.0073	0.0072	0.0071	0.0071	0.0071	0.0072
10000	Chisq(5)	Off by 0.5	10	0.0276	0.0277	0.0278	0.0277	0.0277	0.0277	0.0288	0.0288	0.0289	0.0288	0.0288	0.0288
10000	Chisq(5)	Off by 0.5	20	0.0177	0.0177	0.0177	0.0177	0.0177	0.0176	0.0191	0.0189	0.0189	0.0189	0.0188	0.0188
10000	Chisq(5)	Off by 0.5	40	0.0097	0.0097	0.0097	0.0098	0.0098	0.0098	0.0115	0.0114	0.0113	0.0113	0.0113	0.0113
10000	Chisq(5)	Off by 0.5	60	0.0065	0.0065	0.0065	0.0065	0.0065	0.0065	0.0086	0.0085	0.0084	0.0084	0.0084	0.0084
10000	Chisq(5)	Off by 0.5	80	0.0050	0.0051	0.0051	0.0051	0.0051	0.0051	0.0072	0.0073	0.0072	0.0072	0.0072	0.0072
10000	Chisq(5)	Off by 1.0	10	0.0314	0.0316	0.0315	0.0316	0.0316	0.0316	0.0325	0.0326	0.0325	0.0325	0.0325	0.0325
10000	Chisq(5)	Off by 1.0	20	0.0201	0.0203	0.0202	0.0202	0.0202	0.0202	0.0213	0.0214	0.0213	0.0213	0.0212	0.0212
10000	Chisq(5)	Off by 1.0	40	0.0108	0.0108	0.0109	0.0109	0.0109	0.0109	0.0124	0.0123	0.0123	0.0123	0.0123	0.0123
10000	Chisq(5)	Off by 1.0	60	0.0074	0.0073	0.0074	0.0073	0.0073	0.0073	0.0092	0.0091	0.0091	0.0090	0.0090	0.0090
10000	Chisq(5)	Off by 1.0	80	0.0056	0.0055	0.0055	0.0056	0.0055	0.0055	0.0077	0.0075	0.0075	0.0075	0.0074	0.0075
10000	Chisq(5)	Off by 2.5	10	0.0445	0.0443	0.0444	0.0444	0.0443	0.0444	0.0454	0.0452	0.0452	0.0452	0.0452	0.0452
10000	Chisq(5)	Off by 2.5	20	0.0307	0.0307	0.0308	0.0309	0.0308	0.0308	0.0316	0.0316	0.0316	0.0316	0.0316	0.0316
10000	Chisq(5)	Off by 2.5	40	0.0174	0.0175	0.0174	0.0175	0.0175	0.0174	0.0186	0.0186	0.0185	0.0186	0.0186	0.0185
10000	Chisq(5)	Off by 2.5	60	0.0115	0.0115	0.0115	0.0115	0.0115	0.0115	0.0130	0.0128	0.0128	0.0128	0.0128	0.0128
10000	Chisq(5)	Off by 2.5	80	0.0086	0.0087	0.0087	0.0087	0.0086	0.0086	0.0103	0.0102	0.0102	0.0102	0.0101	0.0101

Table 10 Bias and RMSE for 90th Percentile by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0122	0.0080	0.0065	0.0082	0.0056	0.0071	0.2009	0.1789	0.1784	0.1770	0.1732	0.1710
200	Normal	On-target	20	0.0121	0.0110	0.0146	0.0112	0.0102	0.0108	0.1610	0.1464	0.1423	0.1389	0.1386	0.1368
200	Normal	On-target	40	0.0116	0.0087	0.0092	0.0092	0.0118	0.0114	0.1427	0.1250	0.1242	0.1200	0.1223	0.1202
200	Normal	On-target	60	0.0122	0.0105	0.0117	0.0087	0.0109	0.0092	0.1359	0.1244	0.1212	0.1177	0.1175	0.1171
200	Normal	On-target	80	0.0097	0.0119	0.0086	0.0111	0.0109	0.0099	0.1317	0.1216	0.1170	0.1156	0.1162	0.1155
200	Normal	Off by 0.5	10	-0.0010	-0.0009	-0.0043	-0.0022	-0.0026	-0.0035	0.1928	0.1735	0.1679	0.1679	0.1686	0.1653
200	Normal	Off by 0.5	20	0.0060	0.0062	0.0028	0.0013	0.0017	0.0020	0.1583	0.1437	0.1397	0.1345	0.1363	0.1348
200	Normal	Off by 0.5	40	0.0084	0.0045	0.0026	0.0049	0.0040	0.0035	0.1406	0.1257	0.1214	0.1199	0.1179	0.1165
200	Normal	Off by 0.5	60	0.0042	0.0048	0.0010	0.0024	0.0019	0.0019	0.1360	0.1209	0.1203	0.1179	0.1158	0.1160
200	Normal	Off by 0.5	80	0.0051	0.0017	0.0015	0.0030	0.0032	0.0033	0.1316	0.1197	0.1174	0.1154	0.1146	0.1139
200	Normal	Off by 1.0	10	-0.0132	-0.0146	-0.0141	-0.0136	-0.0140	-0.0128	0.1924	0.1778	0.1769	0.1746	0.1720	0.1735
200	Normal	Off by 1.0	20	-0.0042	-0.0054	-0.0034	-0.0046	-0.0036	-0.0041	0.1565	0.1428	0.1370	0.1349	0.1339	0.1326
200	Normal	Off by 1.0	40	-0.0025	-0.0051	-0.0043	-0.0048	-0.0048	-0.0043	0.1419	0.1269	0.1194	0.1207	0.1195	0.1180
200	Normal	Off by 1.0	60	-0.0060	-0.0066	-0.0067	-0.0065	-0.0070	-0.0075	0.1363	0.1218	0.1218	0.1180	0.1191	0.1179
200	Normal	Off by 1.0	80	-0.0037	-0.0075	-0.0064	-0.0067	-0.0051	-0.0053	0.1315	0.1191	0.1173	0.1149	0.1147	0.1142
200	Uniform	On-target	10	-0.0641	-0.0666	-0.0642	-0.0675	-0.0665	-0.0679	0.2074	0.1942	0.1899	0.1879	0.1869	0.1876
200	Uniform	On-target	20	-0.0499	-0.0529	-0.0527	-0.0532	-0.0524	-0.0536	0.1550	0.1419	0.1385	0.1352	0.1337	0.1336
200	Uniform	On-target	40	-0.0491	-0.0536	-0.0506	-0.0510	-0.0524	-0.0513	0.1317	0.1182	0.1149	0.1115	0.1110	0.1107
200	Uniform	On-target	60	-0.0416	-0.0433	-0.0414	-0.0435	-0.0419	-0.0423	0.1164	0.1073	0.1021	0.1022	0.0992	0.0990
200	Uniform	On-target	80	-0.0378	-0.0386	-0.0370	-0.0380	-0.0379	-0.0378	0.1118	0.1029	0.0980	0.0978	0.0970	0.0959
200	Uniform	Off by 0.5	10	-0.0774	-0.0762	-0.0789	-0.0788	-0.0772	-0.0771	0.2030	0.1904	0.1879	0.1842	0.1836	0.1827
200	Uniform	Off by 0.5	20	-0.0597	-0.0606	-0.0612	-0.0634	-0.0627	-0.0626	0.1581	0.1392	0.1351	0.1331	0.1311	0.1314
200	Uniform	Off by 0.5	40	-0.0578	-0.0574	-0.0573	-0.0580	-0.0590	-0.0584	0.1350	0.1196	0.1148	0.1134	0.1133	0.1136
200	Uniform	Off by 0.5	60	-0.0460	-0.0524	-0.0495	-0.0498	-0.0496	-0.0494	0.1149	0.1076	0.1043	0.1016	0.1004	0.0999
200	Uniform	Off by 0.5	80	-0.0454	-0.0432	-0.0432	-0.0439	-0.0440	-0.0437	0.1161	0.1013	0.1006	0.0987	0.0976	0.0976
200	Uniform	Off by 1.0	10	-0.0926	-0.0838	-0.0840	-0.0846	-0.0844	-0.0849	0.2109	0.1883	0.1850	0.1835	0.1814	0.1812
200	Uniform	Off by 1.0	20	-0.0713	-0.0760	-0.0728	-0.0738	-0.0732	-0.0738	0.1561	0.1475	0.1416	0.1390	0.1394	0.1387
200	Uniform	Off by 1.0	40	-0.0691	-0.0693	-0.0678	-0.0679	-0.0689	-0.0690	0.1345	0.1268	0.1221	0.1197	0.1191	0.1193
200	Uniform	Off by 1.0	60	-0.0566	-0.0580	-0.0581	-0.0581	-0.0591	-0.0590	0.1203	0.1098	0.1056	0.1026	0.1037	0.1029
200	Uniform	Off by 1.0	80	-0.0561	-0.0565	-0.0539	-0.0556	-0.0549	-0.0544	0.1172	0.1073	0.1018	0.1032	0.1014	0.1009
200	Bimodal	On-target	10	0.2479	0.2505	0.2465	0.2475	0.2480	0.2470	0.3092	0.2993	0.2939	0.2936	0.2937	0.2926
200	Bimodal	On-target	20	0.2677	0.2693	0.2683	0.2661	0.2668	0.2675	0.3070	0.2993	0.2951	0.2923	0.2924	0.2936
200	Bimodal	On-target	40	0.2616	0.2613	0.2601	0.2609	0.2613	0.2606	0.2848	0.2793	0.2766	0.2767	0.2770	0.2760
200	Bimodal	On-target	60	0.2592	0.2592	0.2603	0.2612	0.2604	0.2607	0.2779	0.2736	0.2735	0.2741	0.2729	0.2729
200	Bimodal	On-target	80	0.2531	0.2550	0.2532	0.2534	0.2534	0.2539	0.2706	0.2679	0.2651	0.2652	0.2649	0.2651
200	Bimodal	Off by 0.5	10	0.2371	0.2355	0.2346	0.2352	0.2347	0.2355	0.2923	0.2845	0.2802	0.2789	0.2774	0.2781
200	Bimodal	Off by 0.5	20	0.2547	0.2571	0.2547	0.2553	0.2528	0.2558	0.2905	0.2850	0.2809	0.2804	0.2774	0.2804
200	Bimodal	Off by 0.5	40	0.2519	0.2512	0.2511	0.2524	0.2518	0.2513	0.2771	0.2696	0.2676	0.2687	0.2675	0.2668
200	Bimodal	Off by 0.5	60	0.2531	0.2509	0.2517	0.2501	0.2516	0.2513	0.2727	0.2652	0.2651	0.2629	0.2642	0.2639
200	Bimodal	Off by 0.5	80	0.2454	0.2474	0.2452	0.2451	0.2461	0.2454	0.2625	0.2611	0.2569	0.2569	0.2575	0.2568
200	Bimodal	Off by 1.0	10	0.2287	0.2278	0.2275	0.2256	0.2261	0.2262	0.2886	0.2763	0.2742	0.2703	0.2696	0.2698
200	Bimodal	Off by 1.0	20	0.2406	0.2420	0.2401	0.2420	0.2420	0.2428	0.2792	0.2734	0.2691	0.2687	0.2688	0.2691

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 1.0	40	0.2430	0.2433	0.2409	0.2415	0.2409	0.2411	0.2687	0.2608	0.2578	0.2579	0.2567	0.2570
200	Bimodal	Off by 1.0	60	0.2455	0.2417	0.2422	0.2423	0.2423	0.2422	0.2654	0.2564	0.2564	0.2553	0.2554	0.2551
200	Bimodal	Off by 1.0	80	0.2365	0.2381	0.2374	0.2381	0.2371	0.2369	0.2533	0.2517	0.2500	0.2498	0.2487	0.2488
200	Chisq(5)	On-target	10	-0.0712	-0.0756	-0.0769	-0.0763	-0.0763	-0.0741	0.2306	0.2235	0.2205	0.2172	0.2158	0.2152
200	Chisq(5)	On-target	20	-0.0547	-0.0528	-0.0539	-0.0539	-0.0538	-0.0542	0.2069	0.1872	0.1849	0.1839	0.1811	0.1828
200	Chisq(5)	On-target	40	-0.0327	-0.0312	-0.0315	-0.0298	-0.0316	-0.0298	0.1947	0.1768	0.1751	0.1725	0.1712	0.1719
200	Chisq(5)	On-target	60	-0.0125	-0.0148	-0.0149	-0.0152	-0.0154	-0.0156	0.1975	0.1815	0.1785	0.1768	0.1750	0.1761
200	Chisq(5)	On-target	80	-0.0110	-0.0139	-0.0116	-0.0084	-0.0107	-0.0101	0.1939	0.1794	0.1760	0.1762	0.1735	0.1740
200	Chisq(5)	Off by 0.5	10	-0.0608	-0.0568	-0.0541	-0.0559	-0.0566	-0.0558	0.2262	0.2149	0.2132	0.2112	0.2100	0.2094
200	Chisq(5)	Off by 0.5	20	-0.0415	-0.0435	-0.0429	-0.0443	-0.0438	-0.0424	0.1985	0.1819	0.1794	0.1785	0.1751	0.1761
200	Chisq(5)	Off by 0.5	40	-0.0307	-0.0305	-0.0285	-0.0299	-0.0301	-0.0303	0.1909	0.1796	0.1752	0.1722	0.1718	0.1721
200	Chisq(5)	Off by 0.5	60	-0.0122	-0.0171	-0.0183	-0.0164	-0.0166	-0.0166	0.1929	0.1808	0.1785	0.1763	0.1771	0.1756
200	Chisq(5)	Off by 0.5	80	-0.0120	-0.0134	-0.0142	-0.0123	-0.0126	-0.0124	0.1936	0.1818	0.1799	0.1790	0.1769	0.1772
200	Chisq(5)	Off by 1.0	10	-0.0435	-0.0445	-0.0458	-0.0427	-0.0414	-0.0431	0.2267	0.2108	0.2059	0.2031	0.2034	0.2025
200	Chisq(5)	Off by 1.0	20	-0.0356	-0.0407	-0.0399	-0.0387	-0.0384	-0.0391	0.1987	0.1877	0.1806	0.1813	0.1766	0.1774
200	Chisq(5)	Off by 1.0	40	-0.0335	-0.0315	-0.0322	-0.0319	-0.0314	-0.0315	0.1912	0.1794	0.1777	0.1753	0.1748	0.1736
200	Chisq(5)	Off by 1.0	60	-0.0179	-0.0196	-0.0189	-0.0200	-0.0207	-0.0199	0.1971	0.1823	0.1793	0.1780	0.1767	0.1760
200	Chisq(5)	Off by 1.0	80	-0.0197	-0.0194	-0.0162	-0.0191	-0.0184	-0.0186	0.1923	0.1818	0.1801	0.1797	0.1782	0.1785

Table 11 Bias and RMSE for 90th Percentile by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0040	0.0049	0.0042	0.0039	0.0040	0.0038	0.0670	0.0601	0.0596	0.0582	0.0582	0.0578
2000	Normal	On-target	20	0.0004	0.0015	0.0016	0.0014	0.0016	0.0013	0.0532	0.0474	0.0468	0.0465	0.0461	0.0455
2000	Normal	On-target	40	-0.0006	0.0012	0.0007	0.0011	0.0008	0.0011	0.0496	0.0444	0.0437	0.0428	0.0423	0.0428
2000	Normal	On-target	60	0.0031	0.0030	0.0028	0.0027	0.0029	0.0023	0.0461	0.0430	0.0422	0.0412	0.0412	0.0411
2000	Normal	On-target	80	0.0033	0.0027	0.0032	0.0031	0.0031	0.0027	0.0454	0.0409	0.0406	0.0398	0.0402	0.0396
2000	Normal	Off by 0.5	10	0.0033	0.0016	0.0010	0.0009	0.0012	0.0013	0.0616	0.0555	0.0538	0.0537	0.0531	0.0529
2000	Normal	Off by 0.5	20	-0.0006	0.0009	-0.0001	-0.0006	-0.0005	-0.0005	0.0558	0.0505	0.0495	0.0487	0.0485	0.0479
2000	Normal	Off by 0.5	40	0.0015	0.0003	0.0008	0.0009	0.0008	0.0010	0.0481	0.0436	0.0430	0.0416	0.0415	0.0415
2000	Normal	Off by 0.5	60	-0.0005	0.0000	0.0004	0.0003	0.0005	0.0002	0.0466	0.0413	0.0405	0.0400	0.0396	0.0398
2000	Normal	Off by 0.5	80	0.0028	0.0033	0.0025	0.0029	0.0029	0.0030	0.0463	0.0418	0.0414	0.0410	0.0409	0.0405
2000	Normal	Off by 1.0	10	-0.0005	-0.0004	0.0007	-0.0003	-0.0001	-0.0001	0.0649	0.0575	0.0561	0.0561	0.0557	0.0549
2000	Normal	Off by 1.0	20	-0.0023	-0.0014	-0.0004	-0.0004	-0.0007	-0.0010	0.0550	0.0480	0.0473	0.0466	0.0459	0.0464
2000	Normal	Off by 1.0	40	-0.0010	-0.0023	-0.0014	-0.0010	-0.0012	-0.0013	0.0470	0.0436	0.0419	0.0418	0.0413	0.0410
2000	Normal	Off by 1.0	60	0.0016	0.0006	0.0011	0.0005	0.0004	0.0006	0.0453	0.0399	0.0391	0.0387	0.0384	0.0387
2000	Normal	Off by 1.0	80	-0.0025	-0.0025	-0.0018	-0.0017	-0.0015	-0.0018	0.0460	0.0418	0.0408	0.0404	0.0402	0.0402
2000	Normal	Off by 2.5	10	-0.0043	-0.0066	-0.0066	-0.0063	-0.0068	-0.0063	0.0724	0.0659	0.0655	0.0660	0.0653	0.0650
2000	Normal	Off by 2.5	20	-0.0078	-0.0057	-0.0059	-0.0055	-0.0055	-0.0058	0.0589	0.0533	0.0532	0.0522	0.0519	0.0514
2000	Normal	Off by 2.5	40	-0.0086	-0.0076	-0.0076	-0.0075	-0.0075	-0.0076	0.0494	0.0454	0.0440	0.0442	0.0435	0.0435
2000	Normal	Off by 2.5	60	-0.0071	-0.0057	-0.0058	-0.0062	-0.0056	-0.0059	0.0469	0.0428	0.0429	0.0417	0.0419	0.0417
2000	Normal	Off by 2.5	80	-0.0083	-0.0076	-0.0074	-0.0074	-0.0078	-0.0080	0.0449	0.0417	0.0410	0.0408	0.0405	0.0405
2000	Uniform	On-target	10	-0.0798	-0.0818	-0.0810	-0.0811	-0.0812	-0.0811	0.1019	0.1000	0.0988	0.0985	0.0984	0.0983
2000	Uniform	On-target	20	-0.0686	-0.0685	-0.0686	-0.0683	-0.0688	-0.0682	0.0858	0.0827	0.0827	0.0821	0.0820	0.0817
2000	Uniform	On-target	40	-0.0598	-0.0596	-0.0590	-0.0588	-0.0592	-0.0590	0.0732	0.0712	0.0703	0.0695	0.0698	0.0695
2000	Uniform	On-target	60	-0.0519	-0.0524	-0.0516	-0.0520	-0.0520	-0.0518	0.0655	0.0635	0.0624	0.0621	0.0622	0.0619
2000	Uniform	On-target	80	-0.0451	-0.0451	-0.0451	-0.0451	-0.0452	-0.0452	0.0583	0.0554	0.0553	0.0552	0.0552	0.0550
2000	Uniform	Off by 0.5	10	-0.0794	-0.0799	-0.0803	-0.0800	-0.0804	-0.0804	0.1012	0.0985	0.0976	0.0969	0.0969	0.0972
2000	Uniform	Off by 0.5	20	-0.0735	-0.0730	-0.0724	-0.0724	-0.0730	-0.0728	0.0897	0.0854	0.0845	0.0840	0.0844	0.0843
2000	Uniform	Off by 0.5	40	-0.0616	-0.0608	-0.0611	-0.0614	-0.0609	-0.0613	0.0750	0.0713	0.0710	0.0709	0.0703	0.0704
2000	Uniform	Off by 0.5	60	-0.0518	-0.0519	-0.0522	-0.0516	-0.0516	-0.0519	0.0646	0.0622	0.0621	0.0613	0.0612	0.0613
2000	Uniform	Off by 0.5	80	-0.0482	-0.0480	-0.0489	-0.0490	-0.0487	-0.0487	0.0594	0.0585	0.0587	0.0585	0.0580	0.0580
2000	Uniform	Off by 1.0	10	-0.0826	-0.0832	-0.0826	-0.0825	-0.0826	-0.0825	0.1026	0.0994	0.0984	0.0976	0.0979	0.0980
2000	Uniform	Off by 1.0	20	-0.0737	-0.0720	-0.0729	-0.0728	-0.0726	-0.0723	0.0905	0.0856	0.0857	0.0851	0.0849	0.0845
2000	Uniform	Off by 1.0	40	-0.0609	-0.0612	-0.0607	-0.0612	-0.0609	-0.0608	0.0738	0.0717	0.0707	0.0708	0.0704	0.0702
2000	Uniform	Off by 1.0	60	-0.0539	-0.0537	-0.0539	-0.0539	-0.0540	-0.0539	0.0660	0.0636	0.0633	0.0630	0.0630	0.0626
2000	Uniform	Off by 1.0	80	-0.0489	-0.0490	-0.0489	-0.0489	-0.0488	-0.0488	0.0604	0.0594	0.0584	0.0580	0.0579	0.0579
2000	Uniform	Off by 2.5	10	-0.1062	-0.1059	-0.1062	-0.1056	-0.1054	-0.1059	0.1260	0.1231	0.1225	0.1215	0.1212	0.1215
2000	Uniform	Off by 2.5	20	-0.0921	-0.0914	-0.0912	-0.0909	-0.0913	-0.0913	0.1064	0.1038	0.1031	0.1026	0.1028	0.1030
2000	Uniform	Off by 2.5	40	-0.0726	-0.0733	-0.0725	-0.0727	-0.0725	-0.0726	0.0852	0.0843	0.0827	0.0827	0.0825	0.0825
2000	Uniform	Off by 2.5	60	-0.0626	-0.0634	-0.0633	-0.0637	-0.0635	-0.0638	0.0741	0.0727	0.0723	0.0722	0.0720	0.0721
2000	Uniform	Off by 2.5	80	-0.0608	-0.0604	-0.0603	-0.0600	-0.0603	-0.0605	0.0714	0.0692	0.0687	0.0680	0.0683	0.0683
2000	Bimodal	On-target	10	0.2499	0.2486	0.2484	0.2490	0.2484	0.2482	0.2578	0.2549	0.2543	0.2548	0.2541	0.2538
2000	Bimodal	On-target	20	0.2580	0.2574	0.2582	0.2577	0.2577	0.2577	0.2624	0.2613	0.2618	0.2612	0.2610	0.2611

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	0.2567	0.2566	0.2565	0.2560	0.2565	0.2562	0.2600	0.2591	0.2589	0.2583	0.2587	0.2585
2000	Bimodal	On-target	60	0.2512	0.2505	0.2502	0.2505	0.2507	0.2507	0.2539	0.2527	0.2523	0.2526	0.2527	0.2527
2000	Bimodal	On-target	80	0.2460	0.2467	0.2464	0.2465	0.2465	0.2465	0.2484	0.2485	0.2482	0.2482	0.2481	0.2482
2000	Bimodal	Off by 0.5	10	0.2469	0.2451	0.2459	0.2453	0.2449	0.2445	0.2544	0.2510	0.2516	0.2508	0.2504	0.2500
2000	Bimodal	Off by 0.5	20	0.2566	0.2547	0.2550	0.2551	0.2550	0.2550	0.2614	0.2583	0.2584	0.2583	0.2582	0.2581
2000	Bimodal	Off by 0.5	40	0.2540	0.2539	0.2545	0.2538	0.2538	0.2538	0.2571	0.2562	0.2567	0.2559	0.2560	0.2559
2000	Bimodal	Off by 0.5	60	0.2510	0.2495	0.2494	0.2494	0.2496	0.2494	0.2535	0.2514	0.2512	0.2511	0.2513	0.2511
2000	Bimodal	Off by 0.5	80	0.2455	0.2448	0.2455	0.2454	0.2454	0.2455	0.2477	0.2464	0.2470	0.2469	0.2469	0.2469
2000	Bimodal	Off by 1.0	10	0.2416	0.2435	0.2436	0.2432	0.2432	0.2433	0.2492	0.2496	0.2495	0.2489	0.2488	0.2489
2000	Bimodal	Off by 1.0	20	0.2502	0.2506	0.2501	0.2505	0.2505	0.2506	0.2549	0.2543	0.2537	0.2539	0.2539	0.2539
2000	Bimodal	Off by 1.0	40	0.2496	0.2495	0.2496	0.2499	0.2500	0.2499	0.2530	0.2520	0.2519	0.2522	0.2523	0.2521
2000	Bimodal	Off by 1.0	60	0.2466	0.2463	0.2460	0.2464	0.2463	0.2464	0.2493	0.2484	0.2480	0.2483	0.2481	0.2482
2000	Bimodal	Off by 1.0	80	0.2427	0.2420	0.2423	0.2421	0.2422	0.2421	0.2451	0.2438	0.2441	0.2438	0.2439	0.2438
2000	Bimodal	Off by 2.5	10	0.2245	0.2243	0.2245	0.2240	0.2234	0.2236	0.2349	0.2332	0.2332	0.2326	0.2319	0.2322
2000	Bimodal	Off by 2.5	20	0.2329	0.2344	0.2342	0.2341	0.2338	0.2342	0.2388	0.2394	0.2386	0.2384	0.2381	0.2385
2000	Bimodal	Off by 2.5	40	0.2420	0.2431	0.2427	0.2431	0.2428	0.2429	0.2461	0.2464	0.2459	0.2462	0.2458	0.2459
2000	Bimodal	Off by 2.5	60	0.2385	0.2377	0.2377	0.2384	0.2381	0.2381	0.2416	0.2402	0.2401	0.2407	0.2403	0.2403
2000	Bimodal	Off by 2.5	80	0.2371	0.2355	0.2357	0.2359	0.2358	0.2360	0.2399	0.2378	0.2378	0.2380	0.2379	0.2380
2000	Chisq(5)	On-target	10	-0.0788	-0.0783	-0.0783	-0.0774	-0.0778	-0.0781	0.1088	0.1034	0.1027	0.1018	0.1020	0.1021
2000	Chisq(5)	On-target	20	-0.0661	-0.0661	-0.0664	-0.0662	-0.0663	-0.0662	0.0936	0.0907	0.0899	0.0889	0.0890	0.0887
2000	Chisq(5)	On-target	40	-0.0417	-0.0399	-0.0405	-0.0396	-0.0402	-0.0399	0.0741	0.0696	0.0692	0.0682	0.0685	0.0678
2000	Chisq(5)	On-target	60	-0.0287	-0.0282	-0.0282	-0.0281	-0.0280	-0.0282	0.0664	0.0638	0.0634	0.0625	0.0618	0.0622
2000	Chisq(5)	On-target	80	-0.0180	-0.0201	-0.0190	-0.0190	-0.0188	-0.0188	0.0629	0.0601	0.0591	0.0583	0.0584	0.0579
2000	Chisq(5)	Off by 0.5	10	-0.0545	-0.0537	-0.0545	-0.0546	-0.0548	-0.0548	0.0904	0.0859	0.0862	0.0854	0.0851	0.0849
2000	Chisq(5)	Off by 0.5	20	-0.0499	-0.0486	-0.0494	-0.0493	-0.0488	-0.0488	0.0826	0.0772	0.0772	0.0764	0.0758	0.0759
2000	Chisq(5)	Off by 0.5	40	-0.0337	-0.0340	-0.0330	-0.0330	-0.0330	-0.0331	0.0696	0.0668	0.0643	0.0636	0.0638	0.0637
2000	Chisq(5)	Off by 0.5	60	-0.0230	-0.0241	-0.0241	-0.0242	-0.0236	-0.0241	0.0679	0.0651	0.0645	0.0630	0.0634	0.0635
2000	Chisq(5)	Off by 0.5	80	-0.0161	-0.0165	-0.0171	-0.0166	-0.0163	-0.0164	0.0623	0.0590	0.0578	0.0570	0.0570	0.0567
2000	Chisq(5)	Off by 1.0	10	-0.0267	-0.0273	-0.0280	-0.0278	-0.0280	-0.0283	0.0779	0.0700	0.0703	0.0688	0.0679	0.0688
2000	Chisq(5)	Off by 1.0	20	-0.0292	-0.0299	-0.0293	-0.0295	-0.0298	-0.0299	0.0729	0.0684	0.0668	0.0662	0.0659	0.0659
2000	Chisq(5)	Off by 1.0	40	-0.0224	-0.0226	-0.0219	-0.0229	-0.0228	-0.0224	0.0675	0.0636	0.0619	0.0619	0.0618	0.0616
2000	Chisq(5)	Off by 1.0	60	-0.0164	-0.0156	-0.0159	-0.0156	-0.0154	-0.0159	0.0648	0.0599	0.0601	0.0596	0.0593	0.0593
2000	Chisq(5)	Off by 1.0	80	-0.0137	-0.0140	-0.0137	-0.0137	-0.0137	-0.0137	0.0633	0.0578	0.0582	0.0576	0.0571	0.0574
2000	Chisq(5)	Off by 2.5	10	0.0523	0.0526	0.0524	0.0525	0.0526	0.0522	0.0953	0.0912	0.0889	0.0895	0.0885	0.0882
2000	Chisq(5)	Off by 2.5	20	0.0200	0.0211	0.0199	0.0194	0.0201	0.0198	0.0726	0.0669	0.0660	0.0654	0.0650	0.0650
2000	Chisq(5)	Off by 2.5	40	-0.0016	0.0004	-0.0008	0.0000	0.0000	-0.0001	0.0622	0.0588	0.0580	0.0565	0.0561	0.0562
2000	Chisq(5)	Off by 2.5	60	-0.0024	-0.0052	-0.0051	-0.0045	-0.0045	-0.0045	0.0620	0.0573	0.0562	0.0562	0.0560	0.0562
2000	Chisq(5)	Off by 2.5	80	-0.0081	-0.0069	-0.0070	-0.0074	-0.0070	-0.0074	0.0622	0.0583	0.0577	0.0577	0.0570	0.0575

Table 12 Bias and RMSE for 90th Percentile by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	-0.0007	-0.0012	-0.0012	-0.0010	-0.0011	-0.0011	0.0355	0.0339	0.0330	0.0331	0.0327	0.0329
10000	Normal	On-target	20	-0.0014	-0.0013	-0.0015	-0.0014	-0.0012	-0.0013	0.0300	0.0287	0.0278	0.0274	0.0274	0.0275
10000	Normal	On-target	40	-0.0001	-0.0003	-0.0002	0.0000	0.0000	-0.0001	0.0283	0.0266	0.0260	0.0257	0.0257	0.0255
10000	Normal	On-target	60	0.0006	-0.0001	0.0002	0.0001	-0.0001	0.0001	0.0258	0.0246	0.0241	0.0238	0.0237	0.0236
10000	Normal	On-target	80	0.0015	0.0013	0.0011	0.0012	0.0011	0.0011	0.0247	0.0235	0.0230	0.0228	0.0227	0.0228
10000	Normal	Off by 0.5	10	0.0000	-0.0008	-0.0006	-0.0004	-0.0004	-0.0004	0.0349	0.0321	0.0320	0.0316	0.0316	0.0315
10000	Normal	Off by 0.5	20	-0.0009	-0.0008	-0.0008	-0.0005	-0.0008	-0.0008	0.0300	0.0282	0.0275	0.0272	0.0272	0.0270
10000	Normal	Off by 0.5	40	-0.0012	-0.0010	-0.0011	-0.0008	-0.0007	-0.0009	0.0284	0.0259	0.0254	0.0252	0.0253	0.0252
10000	Normal	Off by 0.5	60	-0.0001	-0.0001	0.0001	0.0000	0.0000	-0.0001	0.0257	0.0239	0.0234	0.0236	0.0234	0.0232
10000	Normal	Off by 0.5	80	0.0002	0.0002	0.0003	0.0004	0.0004	0.0004	0.0246	0.0235	0.0230	0.0227	0.0229	0.0227
10000	Normal	Off by 1.0	10	-0.0018	-0.0018	-0.0018	-0.0018	-0.0019	-0.0018	0.0349	0.0326	0.0329	0.0323	0.0324	0.0322
10000	Normal	Off by 1.0	20	-0.0011	-0.0004	-0.0011	-0.0007	-0.0008	-0.0007	0.0302	0.0284	0.0275	0.0271	0.0271	0.0268
10000	Normal	Off by 1.0	40	-0.0006	-0.0006	-0.0008	-0.0006	-0.0007	-0.0007	0.0277	0.0262	0.0259	0.0255	0.0253	0.0253
10000	Normal	Off by 1.0	60	-0.0002	-0.0005	-0.0004	-0.0002	-0.0003	-0.0002	0.0251	0.0240	0.0235	0.0234	0.0234	0.0231
10000	Normal	Off by 1.0	80	-0.0007	-0.0003	-0.0004	-0.0003	-0.0002	-0.0003	0.0248	0.0232	0.0230	0.0226	0.0225	0.0225
10000	Normal	Off by 2.5	10	-0.0037	-0.0035	-0.0031	-0.0033	-0.0034	-0.0034	0.0386	0.0358	0.0357	0.0351	0.0354	0.0353
10000	Normal	Off by 2.5	20	-0.0019	-0.0018	-0.0026	-0.0024	-0.0021	-0.0022	0.0319	0.0299	0.0297	0.0291	0.0288	0.0288
10000	Normal	Off by 2.5	40	-0.0029	-0.0026	-0.0023	-0.0022	-0.0023	-0.0025	0.0287	0.0268	0.0262	0.0260	0.0263	0.0260
10000	Normal	Off by 2.5	60	-0.0023	-0.0021	-0.0021	-0.0020	-0.0019	-0.0021	0.0265	0.0254	0.0250	0.0246	0.0247	0.0246
10000	Normal	Off by 2.5	80	-0.0028	-0.0025	-0.0023	-0.0023	-0.0024	-0.0023	0.0245	0.0237	0.0234	0.0229	0.0230	0.0229
10000	Uniform	On-target	10	-0.0802	-0.0810	-0.0804	-0.0806	-0.0806	-0.0807	0.0878	0.0876	0.0870	0.0871	0.0870	0.0871
10000	Uniform	On-target	20	-0.0710	-0.0706	-0.0710	-0.0709	-0.0711	-0.0711	0.0772	0.0760	0.0763	0.0760	0.0762	0.0762
10000	Uniform	On-target	40	-0.0604	-0.0609	-0.0608	-0.0606	-0.0606	-0.0605	0.0657	0.0655	0.0654	0.0652	0.0652	0.0650
10000	Uniform	On-target	60	-0.0532	-0.0531	-0.0534	-0.0531	-0.0531	-0.0533	0.0582	0.0575	0.0577	0.0573	0.0573	0.0575
10000	Uniform	On-target	80	-0.0476	-0.0475	-0.0476	-0.0477	-0.0475	-0.0475	0.0525	0.0519	0.0520	0.0519	0.0517	0.0516
10000	Uniform	Off by 0.5	10	-0.0829	-0.0833	-0.0831	-0.0831	-0.0831	-0.0829	0.0895	0.0893	0.0888	0.0886	0.0885	0.0883
10000	Uniform	Off by 0.5	20	-0.0730	-0.0728	-0.0729	-0.0730	-0.0732	-0.0731	0.0786	0.0777	0.0778	0.0777	0.0779	0.0777
10000	Uniform	Off by 0.5	40	-0.0617	-0.0619	-0.0622	-0.0617	-0.0619	-0.0618	0.0671	0.0665	0.0666	0.0660	0.0662	0.0661
10000	Uniform	Off by 0.5	60	-0.0543	-0.0544	-0.0546	-0.0545	-0.0544	-0.0545	0.0592	0.0586	0.0589	0.0586	0.0586	0.0586
10000	Uniform	Off by 0.5	80	-0.0485	-0.0486	-0.0486	-0.0487	-0.0485	-0.0487	0.0532	0.0527	0.0528	0.0527	0.0526	0.0526
10000	Uniform	Off by 1.0	10	-0.0855	-0.0858	-0.0856	-0.0854	-0.0857	-0.0859	0.0919	0.0916	0.0914	0.0910	0.0912	0.0913
10000	Uniform	Off by 1.0	20	-0.0754	-0.0748	-0.0746	-0.0747	-0.0748	-0.0749	0.0808	0.0797	0.0795	0.0793	0.0795	0.0794
10000	Uniform	Off by 1.0	40	-0.0641	-0.0632	-0.0632	-0.0630	-0.0631	-0.0630	0.0689	0.0674	0.0675	0.0672	0.0673	0.0671
10000	Uniform	Off by 1.0	60	-0.0555	-0.0558	-0.0557	-0.0558	-0.0557	-0.0557	0.0602	0.0601	0.0598	0.0598	0.0597	0.0597
10000	Uniform	Off by 1.0	80	-0.0500	-0.0502	-0.0503	-0.0502	-0.0502	-0.0501	0.0547	0.0544	0.0544	0.0541	0.0541	0.0540
10000	Uniform	Off by 2.5	10	-0.1000	-0.1005	-0.1002	-0.1003	-0.1003	-0.1003	0.1071	0.1068	0.1063	0.1062	0.1062	0.1062
10000	Uniform	Off by 2.5	20	-0.0849	-0.0846	-0.0847	-0.0846	-0.0844	-0.0844	0.0905	0.0893	0.0893	0.0891	0.0889	0.0889
10000	Uniform	Off by 2.5	40	-0.0690	-0.0683	-0.0687	-0.0690	-0.0690	-0.0689	0.0738	0.0728	0.0730	0.0732	0.0732	0.0730
10000	Uniform	Off by 2.5	60	-0.0605	-0.0607	-0.0606	-0.0606	-0.0605	-0.0607	0.0651	0.0648	0.0646	0.0647	0.0644	0.0646
10000	Uniform	Off by 2.5	80	-0.0550	-0.0552	-0.0554	-0.0554	-0.0555	-0.0555	0.0596	0.0592	0.0593	0.0591	0.0593	0.0592
10000	Bimodal	On-target	10	0.2482	0.2483	0.2479	0.2478	0.2479	0.2478	0.2505	0.2504	0.2499	0.2498	0.2499	0.2498
10000	Bimodal	On-target	20	0.2570	0.2559	0.2558	0.2562	0.2560	0.2562	0.2587	0.2574	0.2573	0.2576	0.2574	0.2576

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	0.2553	0.2552	0.2554	0.2551	0.2552	0.2553	0.2566	0.2564	0.2564	0.2562	0.2563	0.2564
10000	Bimodal	On-target	60	0.2505	0.2506	0.2506	0.2503	0.2506	0.2506	0.2516	0.2515	0.2515	0.2512	0.2515	0.2515
10000	Bimodal	On-target	80	0.2458	0.2459	0.2460	0.2459	0.2460	0.2460	0.2467	0.2467	0.2468	0.2466	0.2467	0.2468
10000	Bimodal	Off by 0.5	10	0.2440	0.2440	0.2443	0.2442	0.2442	0.2442	0.2464	0.2461	0.2464	0.2462	0.2462	0.2461
10000	Bimodal	Off by 0.5	20	0.2533	0.2533	0.2532	0.2534	0.2532	0.2532	0.2551	0.2548	0.2546	0.2548	0.2546	0.2546
10000	Bimodal	Off by 0.5	40	0.2522	0.2523	0.2525	0.2523	0.2522	0.2524	0.2535	0.2534	0.2536	0.2534	0.2533	0.2534
10000	Bimodal	Off by 0.5	60	0.2475	0.2477	0.2477	0.2476	0.2476	0.2475	0.2485	0.2486	0.2486	0.2485	0.2485	0.2484
10000	Bimodal	Off by 0.5	80	0.2432	0.2432	0.2435	0.2434	0.2434	0.2435	0.2441	0.2441	0.2443	0.2442	0.2442	0.2443
10000	Bimodal	Off by 1.0	10	0.2426	0.2418	0.2421	0.2421	0.2423	0.2420	0.2450	0.2438	0.2440	0.2440	0.2443	0.2439
10000	Bimodal	Off by 1.0	20	0.2522	0.2506	0.2517	0.2514	0.2514	0.2515	0.2538	0.2521	0.2531	0.2527	0.2528	0.2528
10000	Bimodal	Off by 1.0	40	0.2508	0.2504	0.2507	0.2506	0.2506	0.2505	0.2521	0.2515	0.2518	0.2517	0.2517	0.2515
10000	Bimodal	Off by 1.0	60	0.2461	0.2462	0.2463	0.2461	0.2461	0.2461	0.2471	0.2471	0.2472	0.2470	0.2470	0.2469
10000	Bimodal	Off by 1.0	80	0.2413	0.2413	0.2416	0.2417	0.2415	0.2415	0.2421	0.2421	0.2424	0.2424	0.2423	0.2422
10000	Bimodal	Off by 2.5	10	0.2284	0.2283	0.2280	0.2281	0.2281	0.2284	0.2313	0.2309	0.2306	0.2305	0.2306	0.2309
10000	Bimodal	Off by 2.5	20	0.2411	0.2404	0.2410	0.2410	0.2408	0.2407	0.2431	0.2421	0.2427	0.2427	0.2424	0.2424
10000	Bimodal	Off by 2.5	40	0.2456	0.2458	0.2458	0.2458	0.2455	0.2455	0.2469	0.2469	0.2468	0.2468	0.2466	0.2466
10000	Bimodal	Off by 2.5	60	0.2428	0.2434	0.2429	0.2432	0.2431	0.2431	0.2440	0.2445	0.2439	0.2442	0.2441	0.2440
10000	Bimodal	Off by 2.5	80	0.2394	0.2392	0.2393	0.2393	0.2392	0.2394	0.2404	0.2401	0.2402	0.2401	0.2401	0.2402
10000	Chisq(5)	On-target	10	-0.0799	-0.0797	-0.0798	-0.0794	-0.0796	-0.0796	0.0891	0.0879	0.0878	0.0874	0.0874	0.0876
10000	Chisq(5)	On-target	20	-0.0619	-0.0623	-0.0623	-0.0625	-0.0625	-0.0627	0.0714	0.0709	0.0706	0.0707	0.0705	0.0707
10000	Chisq(5)	On-target	40	-0.0397	-0.0397	-0.0397	-0.0398	-0.0396	-0.0397	0.0522	0.0510	0.0506	0.0507	0.0505	0.0505
10000	Chisq(5)	On-target	60	-0.0288	-0.0279	-0.0280	-0.0278	-0.0283	-0.0280	0.0432	0.0413	0.0412	0.0409	0.0413	0.0409
10000	Chisq(5)	On-target	80	-0.0198	-0.0203	-0.0197	-0.0194	-0.0198	-0.0197	0.0376	0.0363	0.0358	0.0356	0.0356	0.0356
10000	Chisq(5)	Off by 0.5	10	-0.0518	-0.0532	-0.0529	-0.0527	-0.0528	-0.0527	0.0643	0.0647	0.0640	0.0637	0.0635	0.0636
10000	Chisq(5)	Off by 0.5	20	-0.0460	-0.0468	-0.0465	-0.0467	-0.0463	-0.0463	0.0583	0.0575	0.0574	0.0571	0.0567	0.0567
10000	Chisq(5)	Off by 0.5	40	-0.0314	-0.0315	-0.0314	-0.0314	-0.0314	-0.0311	0.0458	0.0446	0.0444	0.0442	0.0441	0.0439
10000	Chisq(5)	Off by 0.5	60	-0.0226	-0.0229	-0.0229	-0.0228	-0.0229	-0.0228	0.0400	0.0382	0.0379	0.0377	0.0378	0.0376
10000	Chisq(5)	Off by 0.5	80	-0.0163	-0.0166	-0.0165	-0.0163	-0.0166	-0.0166	0.0356	0.0345	0.0343	0.0340	0.0341	0.0338
10000	Chisq(5)	Off by 1.0	10	-0.0286	-0.0284	-0.0280	-0.0280	-0.0283	-0.0283	0.0472	0.0463	0.0459	0.0454	0.0456	0.0456
10000	Chisq(5)	Off by 1.0	20	-0.0309	-0.0304	-0.0305	-0.0307	-0.0307	-0.0305	0.0477	0.0456	0.0455	0.0453	0.0451	0.0450
10000	Chisq(5)	Off by 1.0	40	-0.0241	-0.0244	-0.0242	-0.0241	-0.0241	-0.0241	0.0411	0.0400	0.0398	0.0396	0.0395	0.0394
10000	Chisq(5)	Off by 1.0	60	-0.0188	-0.0182	-0.0182	-0.0182	-0.0183	-0.0184	0.0384	0.0359	0.0356	0.0355	0.0354	0.0355
10000	Chisq(5)	Off by 1.0	80	-0.0130	-0.0136	-0.0135	-0.0136	-0.0137	-0.0136	0.0347	0.0331	0.0331	0.0327	0.0326	0.0326
10000	Chisq(5)	Off by 2.5	10	0.0575	0.0582	0.0585	0.0583	0.0580	0.0581	0.0717	0.0710	0.0708	0.0704	0.0700	0.0701
10000	Chisq(5)	Off by 2.5	20	0.0253	0.0248	0.0250	0.0251	0.0250	0.0250	0.0450	0.0432	0.0426	0.0425	0.0424	0.0423
10000	Chisq(5)	Off by 2.5	40	0.0045	0.0042	0.0043	0.0042	0.0041	0.0042	0.0347	0.0331	0.0325	0.0323	0.0321	0.0320
10000	Chisq(5)	Off by 2.5	60	-0.0017	-0.0011	-0.0014	-0.0014	-0.0016	-0.0013	0.0343	0.0323	0.0314	0.0315	0.0314	0.0312
10000	Chisq(5)	Off by 2.5	80	-0.0020	-0.0023	-0.0023	-0.0024	-0.0023	-0.0023	0.0324	0.0310	0.0307	0.0304	0.0304	0.0304

Table 13 Bias and RMSE for 95th Percentile by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0140	0.0077	0.0081	0.0084	0.0053	0.0062	0.2466	0.2182	0.2123	0.2103	0.2084	0.2068
200	Normal	On-target	20	0.0070	0.0149	0.0173	0.0136	0.0118	0.0135	0.1989	0.1791	0.1708	0.1668	0.1652	0.1654
200	Normal	On-target	40	0.0140	0.0127	0.0136	0.0144	0.0138	0.0132	0.1747	0.1540	0.1485	0.1451	0.1474	0.1440
200	Normal	On-target	60	0.0113	0.0128	0.0111	0.0127	0.0127	0.0118	0.1712	0.1524	0.1467	0.1436	0.1426	0.1418
200	Normal	On-target	80	0.0097	0.0122	0.0077	0.0111	0.0107	0.0111	0.1679	0.1467	0.1415	0.1407	0.1400	0.1395
200	Normal	Off by 0.5	10	-0.0018	-0.0015	-0.0036	-0.0026	-0.0010	-0.0026	0.2302	0.2094	0.2040	0.2005	0.2017	0.1986
200	Normal	Off by 0.5	20	0.0030	0.0082	0.0054	0.0042	0.0045	0.0060	0.1956	0.1766	0.1677	0.1623	0.1639	0.1623
200	Normal	Off by 0.5	40	0.0099	0.0073	0.0061	0.0098	0.0082	0.0078	0.1704	0.1519	0.1463	0.1450	0.1440	0.1417
200	Normal	Off by 0.5	60	0.0055	0.0068	0.0042	0.0038	0.0039	0.0044	0.1686	0.1503	0.1479	0.1420	0.1409	0.1401
200	Normal	Off by 0.5	80	0.0046	0.0035	0.0055	0.0049	0.0053	0.0063	0.1608	0.1473	0.1415	0.1395	0.1384	0.1388
200	Normal	Off by 1.0	10	-0.0151	-0.0150	-0.0142	-0.0136	-0.0135	-0.0128	0.2374	0.2128	0.2091	0.2074	0.2054	0.2061
200	Normal	Off by 1.0	20	-0.0097	-0.0010	-0.0003	-0.0017	-0.0013	0.0000	0.1912	0.1749	0.1640	0.1626	0.1616	0.1596
200	Normal	Off by 1.0	40	0.0041	-0.0022	-0.0027	0.0000	-0.0009	0.0006	0.1670	0.1507	0.1469	0.1449	0.1439	0.1425
200	Normal	Off by 1.0	60	-0.0009	-0.0036	-0.0051	-0.0040	-0.0052	-0.0046	0.1669	0.1488	0.1462	0.1420	0.1414	0.1404
200	Normal	Off by 1.0	80	-0.0051	-0.0023	-0.0038	-0.0037	-0.0029	-0.0021	0.1625	0.1442	0.1423	0.1404	0.1378	0.1388
200	Uniform	On-target	10	0.1221	0.1207	0.1243	0.1219	0.1212	0.1217	0.2653	0.2425	0.2445	0.2394	0.2374	0.2381
200	Uniform	On-target	20	0.1067	0.1101	0.1137	0.1125	0.1116	0.1102	0.2064	0.1850	0.1841	0.1782	0.1757	0.1740
200	Uniform	On-target	40	0.0771	0.0754	0.0740	0.0746	0.0741	0.0737	0.1555	0.1359	0.1343	0.1303	0.1275	0.1266
200	Uniform	On-target	60	0.0582	0.0599	0.0605	0.0593	0.0600	0.0615	0.1343	0.1209	0.1156	0.1135	0.1122	0.1128
200	Uniform	On-target	80	0.0417	0.0476	0.0478	0.0487	0.0476	0.0485	0.1207	0.1081	0.1045	0.1033	0.1007	0.1015
200	Uniform	Off by 0.5	10	0.1078	0.1137	0.1112	0.1086	0.1104	0.1099	0.2458	0.2316	0.2290	0.2227	0.2211	0.2203
200	Uniform	Off by 0.5	20	0.1021	0.0988	0.0994	0.0983	0.0986	0.0996	0.1964	0.1723	0.1664	0.1630	0.1608	0.1620
200	Uniform	Off by 0.5	40	0.0641	0.0641	0.0653	0.0654	0.0643	0.0646	0.1498	0.1304	0.1263	0.1230	0.1212	0.1199
200	Uniform	Off by 0.5	60	0.0512	0.0484	0.0481	0.0499	0.0504	0.0496	0.1301	0.1098	0.1075	0.1054	0.1039	0.1028
200	Uniform	Off by 0.5	80	0.0390	0.0373	0.0397	0.0392	0.0379	0.0387	0.1182	0.1030	0.0993	0.0976	0.0955	0.0953
200	Uniform	Off by 1.0	10	0.0930	0.1015	0.1026	0.1026	0.1020	0.1011	0.2381	0.2203	0.2175	0.2155	0.2122	0.2101
200	Uniform	Off by 1.0	20	0.0937	0.0859	0.0868	0.0852	0.0853	0.0856	0.1898	0.1637	0.1596	0.1555	0.1560	0.1532
200	Uniform	Off by 1.0	40	0.0525	0.0508	0.0506	0.0524	0.0511	0.0507	0.1412	0.1257	0.1189	0.1167	0.1137	0.1131
200	Uniform	Off by 1.0	60	0.0359	0.0379	0.0384	0.0389	0.0376	0.0374	0.1227	0.1048	0.0989	0.0956	0.0955	0.0942
200	Uniform	Off by 1.0	80	0.0244	0.0252	0.0258	0.0253	0.0256	0.0271	0.1084	0.0970	0.0917	0.0895	0.0887	0.0882
200	Bimodal	On-target	10	0.5183	0.5220	0.5199	0.5190	0.5196	0.5189	0.5615	0.5560	0.5516	0.5502	0.5501	0.5494
200	Bimodal	On-target	20	0.5097	0.5120	0.5110	0.5093	0.5098	0.5119	0.5384	0.5320	0.5287	0.5266	0.5270	0.5286
200	Bimodal	On-target	40	0.4671	0.4657	0.4641	0.4639	0.4651	0.4640	0.4857	0.4779	0.4752	0.4745	0.4753	0.4741
200	Bimodal	On-target	60	0.4415	0.4379	0.4431	0.4428	0.4414	0.4417	0.4563	0.4490	0.4529	0.4519	0.4503	0.4507
200	Bimodal	On-target	80	0.4224	0.4230	0.4223	0.4213	0.4206	0.4217	0.4374	0.4328	0.4311	0.4299	0.4289	0.4299
200	Bimodal	Off by 0.5	10	0.5035	0.5005	0.5056	0.5022	0.5059	0.5045	0.5447	0.5329	0.5347	0.5307	0.5336	0.5320
200	Bimodal	Off by 0.5	20	0.4959	0.4981	0.4959	0.4952	0.4932	0.4945	0.5218	0.5173	0.5129	0.5114	0.5090	0.5099
200	Bimodal	Off by 0.5	40	0.4566	0.4531	0.4516	0.4531	0.4526	0.4520	0.4746	0.4652	0.4629	0.4640	0.4628	0.4622
200	Bimodal	Off by 0.5	60	0.4292	0.4291	0.4310	0.4281	0.4292	0.4292	0.4441	0.4400	0.4411	0.4373	0.4382	0.4380
200	Bimodal	Off by 0.5	80	0.4112	0.4118	0.4110	0.4098	0.4099	0.4097	0.4250	0.4218	0.4200	0.4182	0.4179	0.4177
200	Bimodal	Off by 1.0	10	0.4973	0.4968	0.4948	0.4956	0.4956	0.4956	0.5375	0.5282	0.5245	0.5247	0.5236	0.5233
200	Bimodal	Off by 1.0	20	0.4846	0.4822	0.4806	0.4811	0.4806	0.4816	0.5109	0.5023	0.4990	0.4982	0.4976	0.4980

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 1.0	40	0.4409	0.4424	0.4423	0.4397	0.4394	0.4401	0.4597	0.4540	0.4534	0.4500	0.4494	0.4500
200	Bimodal	Off by 1.0	60	0.4227	0.4193	0.4207	0.4200	0.4196	0.4200	0.4384	0.4302	0.4310	0.4293	0.4289	0.4289
200	Bimodal	Off by 1.0	80	0.3998	0.4008	0.3984	0.4007	0.4011	0.3999	0.4142	0.4108	0.4074	0.4090	0.4096	0.4082
200	Chisq(5)	On-target	10	-0.2616	-0.2644	-0.2708	-0.2686	-0.2682	-0.2672	0.3780	0.3704	0.3725	0.3664	0.3651	0.3647
200	Chisq(5)	On-target	20	-0.2010	-0.1932	-0.1974	-0.1974	-0.1974	-0.1968	0.3268	0.2959	0.2968	0.2970	0.2928	0.2943
200	Chisq(5)	On-target	40	-0.1271	-0.1237	-0.1228	-0.1238	-0.1251	-0.1228	0.2728	0.2541	0.2547	0.2504	0.2488	0.2498
200	Chisq(5)	On-target	60	-0.0914	-0.0859	-0.0806	-0.0833	-0.0845	-0.0834	0.2637	0.2451	0.2409	0.2386	0.2388	0.2383
200	Chisq(5)	On-target	80	-0.0595	-0.0653	-0.0619	-0.0608	-0.0616	-0.0608	0.2541	0.2403	0.2368	0.2347	0.2331	0.2328
200	Chisq(5)	Off by 0.5	10	-0.2366	-0.2363	-0.2346	-0.2373	-0.2376	-0.2368	0.3667	0.3496	0.3470	0.3474	0.3458	0.3445
200	Chisq(5)	Off by 0.5	20	-0.1749	-0.1710	-0.1744	-0.1765	-0.1746	-0.1744	0.3054	0.2855	0.2813	0.2794	0.2757	0.2766
200	Chisq(5)	Off by 0.5	40	-0.1207	-0.1162	-0.1136	-0.1173	-0.1173	-0.1143	0.2781	0.2537	0.2501	0.2465	0.2479	0.2457
200	Chisq(5)	Off by 0.5	60	-0.0796	-0.0784	-0.0771	-0.0748	-0.0782	-0.0752	0.2645	0.2452	0.2448	0.2418	0.2404	0.2390
200	Chisq(5)	Off by 0.5	80	-0.0547	-0.0570	-0.0601	-0.0601	-0.0597	-0.0601	0.2516	0.2422	0.2408	0.2354	0.2340	0.2350
200	Chisq(5)	Off by 1.0	10	-0.2092	-0.2139	-0.2119	-0.2108	-0.2106	-0.2126	0.3499	0.3334	0.3291	0.3273	0.3262	0.3266
200	Chisq(5)	Off by 1.0	20	-0.1667	-0.1606	-0.1583	-0.1610	-0.1617	-0.1604	0.3017	0.2841	0.2780	0.2742	0.2731	0.2744
200	Chisq(5)	Off by 1.0	40	-0.1035	-0.1068	-0.1081	-0.1090	-0.1094	-0.1067	0.2701	0.2529	0.2506	0.2487	0.2484	0.2461
200	Chisq(5)	Off by 1.0	60	-0.0720	-0.0756	-0.0723	-0.0741	-0.0737	-0.0734	0.2639	0.2470	0.2410	0.2404	0.2409	0.2400
200	Chisq(5)	Off by 1.0	80	-0.0603	-0.0594	-0.0621	-0.0605	-0.0614	-0.0620	0.2552	0.2448	0.2399	0.2388	0.2373	0.2394

Table 14 Bias and RMSE for 95th Percentile by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0018	0.0043	0.0037	0.0035	0.0035	0.0033	0.0810	0.0731	0.0725	0.0700	0.0705	0.0698
2000	Normal	On-target	20	0.0009	0.0014	0.0014	0.0006	0.0010	0.0004	0.0682	0.0592	0.0580	0.0573	0.0571	0.0562
2000	Normal	On-target	40	0.0036	0.0021	0.0008	0.0011	0.0009	0.0009	0.0619	0.0547	0.0530	0.0523	0.0517	0.0517
2000	Normal	On-target	60	0.0022	0.0046	0.0030	0.0036	0.0026	0.0026	0.0584	0.0524	0.0504	0.0498	0.0496	0.0492
2000	Normal	On-target	80	0.0033	0.0023	0.0026	0.0024	0.0016	0.0020	0.0559	0.0504	0.0493	0.0483	0.0482	0.0479
2000	Normal	Off by 0.5	10	0.0005	-0.0001	0.0004	0.0002	-0.0001	0.0003	0.0770	0.0677	0.0667	0.0660	0.0647	0.0645
2000	Normal	Off by 0.5	20	0.0004	-0.0013	-0.0016	-0.0012	-0.0014	-0.0017	0.0684	0.0600	0.0586	0.0574	0.0575	0.0572
2000	Normal	Off by 0.5	40	-0.0003	0.0003	-0.0001	0.0007	0.0008	0.0008	0.0591	0.0533	0.0514	0.0511	0.0500	0.0497
2000	Normal	Off by 0.5	60	-0.0002	-0.0013	-0.0003	-0.0008	-0.0002	-0.0005	0.0584	0.0508	0.0500	0.0487	0.0486	0.0489
2000	Normal	Off by 0.5	80	0.0041	0.0032	0.0028	0.0031	0.0030	0.0030	0.0566	0.0519	0.0510	0.0493	0.0492	0.0490
2000	Normal	Off by 1.0	10	-0.0014	-0.0019	0.0000	-0.0007	-0.0005	-0.0011	0.0805	0.0712	0.0702	0.0684	0.0684	0.0681
2000	Normal	Off by 1.0	20	-0.0010	-0.0019	-0.0021	-0.0017	-0.0012	-0.0019	0.0667	0.0589	0.0578	0.0555	0.0549	0.0550
2000	Normal	Off by 1.0	40	-0.0006	-0.0027	-0.0021	-0.0015	-0.0014	-0.0014	0.0601	0.0543	0.0518	0.0512	0.0508	0.0505
2000	Normal	Off by 1.0	60	0.0006	0.0011	0.0000	0.0002	0.0003	0.0004	0.0556	0.0500	0.0495	0.0479	0.0476	0.0471
2000	Normal	Off by 1.0	80	-0.0021	-0.0005	-0.0011	-0.0013	-0.0009	-0.0003	0.0547	0.0499	0.0486	0.0479	0.0473	0.0475
2000	Normal	Off by 2.5	10	-0.0048	-0.0071	-0.0074	-0.0074	-0.0072	-0.0068	0.0874	0.0796	0.0787	0.0770	0.0762	0.0762
2000	Normal	Off by 2.5	20	-0.0081	-0.0076	-0.0058	-0.0065	-0.0066	-0.0074	0.0708	0.0637	0.0620	0.0608	0.0610	0.0607
2000	Normal	Off by 2.5	40	-0.0079	-0.0068	-0.0072	-0.0083	-0.0075	-0.0082	0.0604	0.0543	0.0522	0.0518	0.0517	0.0515
2000	Normal	Off by 2.5	60	-0.0065	-0.0074	-0.0076	-0.0077	-0.0067	-0.0072	0.0572	0.0510	0.0500	0.0492	0.0490	0.0492
2000	Normal	Off by 2.5	80	-0.0092	-0.0087	-0.0082	-0.0087	-0.0086	-0.0092	0.0565	0.0516	0.0498	0.0491	0.0488	0.0487
2000	Uniform	On-target	10	0.1025	0.1026	0.1041	0.1040	0.1036	0.1037	0.1286	0.1233	0.1236	0.1235	0.1228	0.1229
2000	Uniform	On-target	20	0.0929	0.0947	0.0933	0.0938	0.0930	0.0935	0.1114	0.1090	0.1071	0.1069	0.1057	0.1060
2000	Uniform	On-target	40	0.0651	0.0645	0.0648	0.0655	0.0655	0.0654	0.0816	0.0773	0.0771	0.0769	0.0767	0.0766
2000	Uniform	On-target	60	0.0492	0.0487	0.0498	0.0490	0.0492	0.0491	0.0667	0.0626	0.0625	0.0611	0.0610	0.0609
2000	Uniform	On-target	80	0.0399	0.0397	0.0401	0.0403	0.0400	0.0400	0.0565	0.0525	0.0521	0.0523	0.0515	0.0517
2000	Uniform	Off by 0.5	10	0.1029	0.1046	0.1035	0.1039	0.1044	0.1039	0.1292	0.1260	0.1233	0.1229	0.1233	0.1228
2000	Uniform	Off by 0.5	20	0.0861	0.0850	0.0853	0.0854	0.0856	0.0855	0.1060	0.1005	0.0994	0.0989	0.0988	0.0986
2000	Uniform	Off by 0.5	40	0.0579	0.0582	0.0590	0.0582	0.0586	0.0584	0.0754	0.0717	0.0715	0.0701	0.0700	0.0697
2000	Uniform	Off by 0.5	60	0.0463	0.0460	0.0452	0.0455	0.0452	0.0449	0.0639	0.0588	0.0573	0.0572	0.0567	0.0565
2000	Uniform	Off by 0.5	80	0.0336	0.0336	0.0327	0.0325	0.0326	0.0325	0.0514	0.0480	0.0470	0.0464	0.0460	0.0458
2000	Uniform	Off by 1.0	10	0.1005	0.1012	0.1010	0.1012	0.1011	0.1007	0.1258	0.1213	0.1199	0.1190	0.1190	0.1189
2000	Uniform	Off by 1.0	20	0.0836	0.0848	0.0845	0.0842	0.0843	0.0847	0.1033	0.0996	0.0985	0.0978	0.0976	0.0979
2000	Uniform	Off by 1.0	40	0.0566	0.0585	0.0573	0.0575	0.0580	0.0580	0.0738	0.0715	0.0694	0.0691	0.0694	0.0693
2000	Uniform	Off by 1.0	60	0.0417	0.0403	0.0407	0.0407	0.0407	0.0408	0.0587	0.0541	0.0535	0.0529	0.0526	0.0527
2000	Uniform	Off by 1.0	80	0.0312	0.0300	0.0301	0.0303	0.0308	0.0308	0.0493	0.0454	0.0445	0.0438	0.0444	0.0438
2000	Uniform	Off by 2.5	10	0.0752	0.0751	0.0751	0.0756	0.0758	0.0754	0.1076	0.1041	0.1039	0.1024	0.1027	0.1019
2000	Uniform	Off by 2.5	20	0.0667	0.0669	0.0672	0.0670	0.0671	0.0668	0.0899	0.0869	0.0870	0.0855	0.0855	0.0853
2000	Uniform	Off by 2.5	40	0.0471	0.0462	0.0475	0.0473	0.0481	0.0473	0.0678	0.0644	0.0640	0.0628	0.0635	0.0626
2000	Uniform	Off by 2.5	60	0.0332	0.0341	0.0337	0.0335	0.0336	0.0332	0.0548	0.0515	0.0507	0.0494	0.0494	0.0490
2000	Uniform	Off by 2.5	80	0.0200	0.0199	0.0209	0.0204	0.0207	0.0205	0.0448	0.0401	0.0393	0.0387	0.0387	0.0381
2000	Bimodal	On-target	10	0.5195	0.5167	0.5174	0.5173	0.5175	0.5173	0.5249	0.5210	0.5215	0.5212	0.5214	0.5211
2000	Bimodal	On-target	20	0.4987	0.4997	0.4998	0.4995	0.4991	0.4995	0.5020	0.5024	0.5023	0.5018	0.5014	0.5018

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	0.4572	0.4566	0.4569	0.4566	0.4569	0.4571	0.4597	0.4585	0.4585	0.4582	0.4585	0.4586
2000	Bimodal	On-target	60	0.4293	0.4296	0.4297	0.4291	0.4298	0.4298	0.4315	0.4312	0.4313	0.4305	0.4312	0.4312
2000	Bimodal	On-target	80	0.4116	0.4127	0.4120	0.4122	0.4123	0.4121	0.4135	0.4141	0.4134	0.4135	0.4136	0.4134
2000	Bimodal	Off by 0.5	10	0.5134	0.5111	0.5118	0.5124	0.5119	0.5114	0.5186	0.5151	0.5155	0.5160	0.5155	0.5150
2000	Bimodal	Off by 0.5	20	0.4937	0.4938	0.4928	0.4937	0.4931	0.4933	0.4973	0.4964	0.4952	0.4960	0.4953	0.4955
2000	Bimodal	Off by 0.5	40	0.4510	0.4513	0.4527	0.4515	0.4519	0.4512	0.4534	0.4530	0.4544	0.4530	0.4534	0.4527
2000	Bimodal	Off by 0.5	60	0.4253	0.4254	0.4248	0.4248	0.4251	0.4248	0.4273	0.4269	0.4262	0.4261	0.4263	0.4260
2000	Bimodal	Off by 0.5	80	0.4088	0.4080	0.4081	0.4077	0.4080	0.4080	0.4105	0.4093	0.4093	0.4089	0.4091	0.4091
2000	Bimodal	Off by 1.0	10	0.5088	0.5107	0.5108	0.5103	0.5105	0.5101	0.5141	0.5149	0.5148	0.5142	0.5143	0.5139
2000	Bimodal	Off by 1.0	20	0.4883	0.4880	0.4866	0.4867	0.4871	0.4872	0.4918	0.4905	0.4890	0.4889	0.4893	0.4894
2000	Bimodal	Off by 1.0	40	0.4450	0.4462	0.4468	0.4461	0.4455	0.4457	0.4475	0.4479	0.4484	0.4477	0.4470	0.4472
2000	Bimodal	Off by 1.0	60	0.4197	0.4189	0.4191	0.4199	0.4198	0.4195	0.4217	0.4204	0.4205	0.4211	0.4211	0.4207
2000	Bimodal	Off by 1.0	80	0.4055	0.4045	0.4045	0.4041	0.4041	0.4044	0.4074	0.4058	0.4057	0.4054	0.4053	0.4056
2000	Bimodal	Off by 2.5	10	0.4909	0.4895	0.4890	0.4898	0.4892	0.4892	0.4975	0.4951	0.4943	0.4949	0.4943	0.4942
2000	Bimodal	Off by 2.5	20	0.4729	0.4731	0.4738	0.4726	0.4728	0.4730	0.4767	0.4762	0.4765	0.4751	0.4754	0.4755
2000	Bimodal	Off by 2.5	40	0.4415	0.4427	0.4426	0.4428	0.4424	0.4424	0.4442	0.4447	0.4445	0.4447	0.4442	0.4441
2000	Bimodal	Off by 2.5	60	0.4151	0.4145	0.4138	0.4155	0.4149	0.4150	0.4174	0.4162	0.4155	0.4171	0.4164	0.4165
2000	Bimodal	Off by 2.5	80	0.3986	0.3993	0.3991	0.3997	0.3991	0.3994	0.4007	0.4008	0.4006	0.4011	0.4004	0.4007
2000	Chisq(5)	On-target	10	-0.2741	-0.2719	-0.2718	-0.2724	-0.2728	-0.2733	0.2892	0.2853	0.2848	0.2850	0.2855	0.2858
2000	Chisq(5)	On-target	20	-0.2110	-0.2123	-0.2124	-0.2119	-0.2122	-0.2122	0.2291	0.2275	0.2269	0.2259	0.2262	0.2258
2000	Chisq(5)	On-target	40	-0.1319	-0.1300	-0.1298	-0.1300	-0.1303	-0.1299	0.1559	0.1514	0.1502	0.1504	0.1504	0.1499
2000	Chisq(5)	On-target	60	-0.0930	-0.0938	-0.0938	-0.0938	-0.0941	-0.0941	0.1227	0.1215	0.1202	0.1196	0.1198	0.1196
2000	Chisq(5)	On-target	80	-0.0658	-0.0661	-0.0661	-0.0662	-0.0673	-0.0665	0.1053	0.1016	0.1005	0.0998	0.1003	0.0997
2000	Chisq(5)	Off by 0.5	10	-0.2398	-0.2370	-0.2371	-0.2380	-0.2380	-0.2379	0.2570	0.2517	0.2512	0.2515	0.2514	0.2511
2000	Chisq(5)	Off by 0.5	20	-0.1806	-0.1808	-0.1812	-0.1814	-0.1805	-0.1807	0.2000	0.1973	0.1975	0.1970	0.1962	0.1964
2000	Chisq(5)	Off by 0.5	40	-0.1148	-0.1157	-0.1158	-0.1156	-0.1154	-0.1157	0.1429	0.1386	0.1383	0.1374	0.1370	0.1372
2000	Chisq(5)	Off by 0.5	60	-0.0816	-0.0832	-0.0825	-0.0830	-0.0830	-0.0828	0.1168	0.1150	0.1139	0.1136	0.1131	0.1134
2000	Chisq(5)	Off by 0.5	80	-0.0638	-0.0613	-0.0605	-0.0609	-0.0609	-0.0610	0.1028	0.0974	0.0967	0.0954	0.0957	0.0956
2000	Chisq(5)	Off by 1.0	10	-0.1989	-0.2006	-0.2002	-0.2005	-0.2000	-0.2009	0.2194	0.2178	0.2164	0.2162	0.2156	0.2165
2000	Chisq(5)	Off by 1.0	20	-0.1484	-0.1518	-0.1519	-0.1512	-0.1511	-0.1507	0.1730	0.1725	0.1719	0.1709	0.1705	0.1699
2000	Chisq(5)	Off by 1.0	40	-0.0994	-0.0983	-0.0971	-0.0975	-0.0975	-0.0970	0.1308	0.1265	0.1246	0.1243	0.1242	0.1237
2000	Chisq(5)	Off by 1.0	60	-0.0697	-0.0694	-0.0701	-0.0696	-0.0695	-0.0688	0.1107	0.1058	0.1058	0.1053	0.1043	0.1037
2000	Chisq(5)	Off by 1.0	80	-0.0510	-0.0508	-0.0513	-0.0515	-0.0519	-0.0519	0.1007	0.0941	0.0934	0.0932	0.0933	0.0928
2000	Chisq(5)	Off by 2.5	10	-0.0878	-0.0879	-0.0894	-0.0887	-0.0892	-0.0901	0.1343	0.1280	0.1273	0.1267	0.1262	0.1264
2000	Chisq(5)	Off by 2.5	20	-0.0854	-0.0839	-0.0868	-0.0864	-0.0865	-0.0863	0.1247	0.1178	0.1189	0.1175	0.1175	0.1171
2000	Chisq(5)	Off by 2.5	40	-0.0643	-0.0653	-0.0639	-0.0648	-0.0645	-0.0646	0.1089	0.1035	0.1020	0.1016	0.1009	0.1009
2000	Chisq(5)	Off by 2.5	60	-0.0518	-0.0527	-0.0528	-0.0510	-0.0515	-0.0514	0.1007	0.0975	0.0959	0.0949	0.0953	0.0951
2000	Chisq(5)	Off by 2.5	80	-0.0418	-0.0435	-0.0417	-0.0429	-0.0431	-0.0422	0.0933	0.0895	0.0884	0.0876	0.0873	0.0870

Table 15 Bias and RMSE for 95th Percentile by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	0.0010	0.0000	-0.0003	0.0004	-0.0001	-0.0003	0.0456	0.0432	0.0424	0.0420	0.0416	0.0419
10000	Normal	On-target	20	0.0003	0.0002	-0.0008	-0.0003	0.0001	-0.0001	0.0386	0.0364	0.0352	0.0348	0.0352	0.0348
10000	Normal	On-target	40	0.0010	0.0016	0.0011	0.0015	0.0013	0.0013	0.0356	0.0335	0.0327	0.0322	0.0321	0.0319
10000	Normal	On-target	60	0.0013	0.0010	0.0013	0.0012	0.0010	0.0014	0.0333	0.0311	0.0309	0.0301	0.0300	0.0300
10000	Normal	On-target	80	0.0020	0.0012	0.0014	0.0014	0.0012	0.0014	0.0310	0.0292	0.0287	0.0284	0.0283	0.0282
10000	Normal	Off by 0.5	10	-0.0012	-0.0008	-0.0012	-0.0009	-0.0011	-0.0011	0.0432	0.0405	0.0399	0.0398	0.0396	0.0394
10000	Normal	Off by 0.5	20	-0.0005	0.0004	0.0006	0.0005	0.0003	0.0005	0.0383	0.0360	0.0351	0.0346	0.0345	0.0344
10000	Normal	Off by 0.5	40	0.0005	0.0000	0.0003	0.0003	0.0002	0.0002	0.0345	0.0325	0.0320	0.0315	0.0314	0.0312
10000	Normal	Off by 0.5	60	0.0016	0.0014	0.0013	0.0014	0.0015	0.0013	0.0322	0.0311	0.0301	0.0297	0.0295	0.0296
10000	Normal	Off by 0.5	80	0.0017	0.0014	0.0019	0.0020	0.0020	0.0021	0.0318	0.0288	0.0288	0.0285	0.0283	0.0282
10000	Normal	Off by 1.0	10	-0.0009	-0.0005	-0.0001	-0.0009	-0.0009	-0.0007	0.0449	0.0415	0.0414	0.0405	0.0407	0.0402
10000	Normal	Off by 1.0	20	0.0010	0.0008	0.0004	0.0007	0.0004	0.0006	0.0379	0.0356	0.0349	0.0343	0.0344	0.0341
10000	Normal	Off by 1.0	40	-0.0006	-0.0016	-0.0014	-0.0012	-0.0010	-0.0013	0.0351	0.0322	0.0317	0.0315	0.0314	0.0310
10000	Normal	Off by 1.0	60	0.0013	0.0008	0.0008	0.0009	0.0009	0.0010	0.0320	0.0302	0.0298	0.0291	0.0291	0.0290
10000	Normal	Off by 1.0	80	0.0012	0.0011	0.0010	0.0013	0.0010	0.0009	0.0312	0.0291	0.0287	0.0284	0.0284	0.0282
10000	Normal	Off by 2.5	10	-0.0020	-0.0022	-0.0018	-0.0021	-0.0023	-0.0020	0.0461	0.0427	0.0431	0.0419	0.0421	0.0420
10000	Normal	Off by 2.5	20	-0.0002	-0.0006	-0.0014	-0.0010	-0.0009	-0.0008	0.0400	0.0375	0.0368	0.0364	0.0364	0.0361
10000	Normal	Off by 2.5	40	-0.0013	-0.0019	-0.0009	-0.0012	-0.0015	-0.0015	0.0363	0.0325	0.0319	0.0316	0.0316	0.0315
10000	Normal	Off by 2.5	60	-0.0006	-0.0014	-0.0011	-0.0012	-0.0011	-0.0013	0.0321	0.0300	0.0295	0.0292	0.0290	0.0290
10000	Normal	Off by 2.5	80	-0.0009	-0.0008	-0.0008	-0.0007	-0.0010	-0.0010	0.0307	0.0295	0.0288	0.0286	0.0286	0.0286
10000	Uniform	On-target	10	0.1067	0.1061	0.1064	0.1065	0.1063	0.1062	0.1162	0.1142	0.1144	0.1143	0.1140	0.1139
10000	Uniform	On-target	20	0.0926	0.0924	0.0920	0.0927	0.0922	0.0924	0.0999	0.0989	0.0980	0.0986	0.0981	0.0982
10000	Uniform	On-target	40	0.0651	0.0654	0.0649	0.0649	0.0652	0.0652	0.0720	0.0713	0.0706	0.0706	0.0707	0.0707
10000	Uniform	On-target	60	0.0481	0.0488	0.0485	0.0487	0.0487	0.0486	0.0552	0.0552	0.0548	0.0547	0.0547	0.0546
10000	Uniform	On-target	80	0.0387	0.0388	0.0386	0.0386	0.0386	0.0386	0.0466	0.0456	0.0455	0.0454	0.0451	0.0451
10000	Uniform	Off by 0.5	10	0.1030	0.1026	0.1031	0.1029	0.1030	0.1029	0.1124	0.1107	0.1112	0.1105	0.1108	0.1106
10000	Uniform	Off by 0.5	20	0.0871	0.0869	0.0872	0.0871	0.0869	0.0872	0.0943	0.0930	0.0933	0.0929	0.0927	0.0931
10000	Uniform	Off by 0.5	40	0.0591	0.0597	0.0593	0.0596	0.0596	0.0596	0.0670	0.0661	0.0657	0.0658	0.0657	0.0658
10000	Uniform	Off by 0.5	60	0.0424	0.0427	0.0432	0.0431	0.0432	0.0431	0.0503	0.0497	0.0501	0.0497	0.0498	0.0497
10000	Uniform	Off by 0.5	80	0.0336	0.0336	0.0334	0.0335	0.0335	0.0334	0.0423	0.0411	0.0409	0.0408	0.0407	0.0406
10000	Uniform	Off by 1.0	10	0.0992	0.0995	0.0994	0.0996	0.0994	0.0993	0.1082	0.1072	0.1069	0.1068	0.1067	0.1066
10000	Uniform	Off by 1.0	20	0.0839	0.0832	0.0835	0.0836	0.0836	0.0838	0.0915	0.0897	0.0898	0.0898	0.0898	0.0897
10000	Uniform	Off by 1.0	40	0.0558	0.0565	0.0560	0.0562	0.0565	0.0563	0.0632	0.0625	0.0620	0.0622	0.0624	0.0621
10000	Uniform	Off by 1.0	60	0.0393	0.0397	0.0401	0.0398	0.0400	0.0397	0.0474	0.0469	0.0471	0.0467	0.0467	0.0465
10000	Uniform	Off by 1.0	80	0.0304	0.0294	0.0297	0.0298	0.0297	0.0298	0.0396	0.0376	0.0378	0.0375	0.0376	0.0376
10000	Uniform	Off by 2.5	10	0.0833	0.0831	0.0828	0.0830	0.0831	0.0830	0.0957	0.0937	0.0930	0.0929	0.0931	0.0930
10000	Uniform	Off by 2.5	20	0.0749	0.0758	0.0750	0.0753	0.0752	0.0756	0.0838	0.0836	0.0828	0.0827	0.0827	0.0829
10000	Uniform	Off by 2.5	40	0.0524	0.0531	0.0531	0.0527	0.0529	0.0529	0.0605	0.0600	0.0598	0.0593	0.0594	0.0594
10000	Uniform	Off by 2.5	60	0.0376	0.0375	0.0379	0.0377	0.0376	0.0376	0.0463	0.0452	0.0453	0.0450	0.0447	0.0447
10000	Uniform	Off by 2.5	80	0.0271	0.0268	0.0265	0.0264	0.0263	0.0264	0.0374	0.0362	0.0358	0.0355	0.0353	0.0353
10000	Bimodal	On-target	10	0.5185	0.5191	0.5186	0.5189	0.5190	0.5187	0.5203	0.5207	0.5201	0.5204	0.5205	0.5202
10000	Bimodal	On-target	20	0.4995	0.4992	0.4991	0.4992	0.4993	0.4994	0.5008	0.5004	0.5002	0.5003	0.5004	0.5005

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	0.4577	0.4581	0.4577	0.4578	0.4577	0.4577	0.4587	0.4589	0.4585	0.4586	0.4585	0.4585
10000	Bimodal	On-target	60	0.4308	0.4307	0.4309	0.4309	0.4309	0.4311	0.4317	0.4315	0.4316	0.4316	0.4316	0.4318
10000	Bimodal	On-target	80	0.4117	0.4122	0.4122	0.4124	0.4121	0.4123	0.4125	0.4129	0.4129	0.4130	0.4127	0.4129
10000	Bimodal	Off by 0.5	10	0.5146	0.5129	0.5140	0.5133	0.5136	0.5136	0.5164	0.5145	0.5156	0.5148	0.5151	0.5150
10000	Bimodal	Off by 0.5	20	0.4924	0.4932	0.4933	0.4932	0.4932	0.4933	0.4938	0.4943	0.4943	0.4943	0.4943	0.4944
10000	Bimodal	Off by 0.5	40	0.4505	0.4515	0.4513	0.4509	0.4511	0.4511	0.4515	0.4523	0.4521	0.4517	0.4519	0.4519
10000	Bimodal	Off by 0.5	60	0.4243	0.4243	0.4243	0.4241	0.4242	0.4242	0.4251	0.4250	0.4250	0.4248	0.4249	0.4249
10000	Bimodal	Off by 0.5	80	0.4071	0.4069	0.4070	0.4069	0.4069	0.4069	0.4079	0.4075	0.4077	0.4075	0.4075	0.4075
10000	Bimodal	Off by 1.0	10	0.5112	0.5104	0.5108	0.5111	0.5111	0.5106	0.5130	0.5120	0.5123	0.5125	0.5126	0.5121
10000	Bimodal	Off by 1.0	20	0.4898	0.4893	0.4903	0.4900	0.4899	0.4898	0.4911	0.4905	0.4914	0.4911	0.4909	0.4909
10000	Bimodal	Off by 1.0	40	0.4476	0.4473	0.4480	0.4478	0.4476	0.4476	0.4486	0.4482	0.4488	0.4486	0.4484	0.4484
10000	Bimodal	Off by 1.0	60	0.4212	0.4214	0.4213	0.4214	0.4214	0.4213	0.4220	0.4221	0.4220	0.4220	0.4221	0.4220
10000	Bimodal	Off by 1.0	80	0.4030	0.4036	0.4037	0.4039	0.4036	0.4037	0.4038	0.4042	0.4043	0.4045	0.4043	0.4043
10000	Bimodal	Off by 2.5	10	0.4962	0.4958	0.4956	0.4961	0.4961	0.4958	0.4981	0.4975	0.4973	0.4977	0.4977	0.4974
10000	Bimodal	Off by 2.5	20	0.4822	0.4817	0.4819	0.4823	0.4817	0.4820	0.4836	0.4828	0.4831	0.4835	0.4828	0.4831
10000	Bimodal	Off by 2.5	40	0.4469	0.4469	0.4471	0.4473	0.4469	0.4468	0.4479	0.4478	0.4480	0.4481	0.4477	0.4476
10000	Bimodal	Off by 2.5	60	0.4206	0.4213	0.4208	0.4211	0.4209	0.4210	0.4215	0.4220	0.4215	0.4218	0.4216	0.4217
10000	Bimodal	Off by 2.5	80	0.4034	0.4033	0.4032	0.4031	0.4030	0.4031	0.4042	0.4040	0.4039	0.4038	0.4037	0.4037
10000	Chisq(5)	On-target	10	-0.2735	-0.2731	-0.2730	-0.2732	-0.2735	-0.2734	0.2782	0.2772	0.2772	0.2773	0.2775	0.2774
10000	Chisq(5)	On-target	20	-0.2042	-0.2047	-0.2051	-0.2050	-0.2046	-0.2049	0.2103	0.2100	0.2102	0.2101	0.2096	0.2099
10000	Chisq(5)	On-target	40	-0.1286	-0.1273	-0.1283	-0.1281	-0.1276	-0.1279	0.1371	0.1350	0.1358	0.1355	0.1349	0.1353
10000	Chisq(5)	On-target	60	-0.0902	-0.0898	-0.0894	-0.0900	-0.0902	-0.0901	0.1016	0.1001	0.0993	0.0999	0.1000	0.0997
10000	Chisq(5)	On-target	80	-0.0680	-0.0676	-0.0678	-0.0674	-0.0676	-0.0674	0.0819	0.0801	0.0802	0.0795	0.0798	0.0797
10000	Chisq(5)	Off by 0.5	10	-0.2327	-0.2340	-0.2337	-0.2337	-0.2338	-0.2334	0.2380	0.2391	0.2385	0.2385	0.2385	0.2382
10000	Chisq(5)	Off by 0.5	20	-0.1771	-0.1770	-0.1772	-0.1775	-0.1773	-0.1774	0.1839	0.1830	0.1831	0.1832	0.1830	0.1832
10000	Chisq(5)	Off by 0.5	40	-0.1105	-0.1100	-0.1100	-0.1105	-0.1101	-0.1099	0.1200	0.1187	0.1183	0.1188	0.1185	0.1182
10000	Chisq(5)	Off by 0.5	60	-0.0779	-0.0776	-0.0773	-0.0775	-0.0775	-0.0776	0.0907	0.0891	0.0889	0.0888	0.0886	0.0885
10000	Chisq(5)	Off by 0.5	80	-0.0581	-0.0584	-0.0579	-0.0583	-0.0585	-0.0583	0.0741	0.0731	0.0722	0.0723	0.0724	0.0723
10000	Chisq(5)	Off by 1.0	10	-0.1992	-0.1982	-0.1977	-0.1976	-0.1984	-0.1984	0.2056	0.2040	0.2034	0.2031	0.2038	0.2039
10000	Chisq(5)	Off by 1.0	20	-0.1509	-0.1522	-0.1521	-0.1526	-0.1524	-0.1521	0.1593	0.1591	0.1589	0.1593	0.1591	0.1588
10000	Chisq(5)	Off by 1.0	40	-0.0965	-0.0967	-0.0964	-0.0966	-0.0967	-0.0967	0.1070	0.1065	0.1060	0.1060	0.1059	0.1060
10000	Chisq(5)	Off by 1.0	60	-0.0698	-0.0687	-0.0683	-0.0692	-0.0690	-0.0688	0.0833	0.0815	0.0810	0.0815	0.0811	0.0811
10000	Chisq(5)	Off by 1.0	80	-0.0513	-0.0526	-0.0523	-0.0519	-0.0519	-0.0519	0.0689	0.0678	0.0676	0.0669	0.0669	0.0668
10000	Chisq(5)	Off by 2.5	10	-0.0820	-0.0805	-0.0803	-0.0806	-0.0811	-0.0808	0.0983	0.0954	0.0946	0.0946	0.0949	0.0949
10000	Chisq(5)	Off by 2.5	20	-0.0784	-0.0786	-0.0784	-0.0787	-0.0780	-0.0783	0.0938	0.0921	0.0915	0.0914	0.0909	0.0909
10000	Chisq(5)	Off by 2.5	40	-0.0602	-0.0606	-0.0609	-0.0604	-0.0605	-0.0605	0.0771	0.0754	0.0754	0.0748	0.0748	0.0747
10000	Chisq(5)	Off by 2.5	60	-0.0481	-0.0471	-0.0470	-0.0473	-0.0476	-0.0474	0.0677	0.0651	0.0648	0.0647	0.0647	0.0645
10000	Chisq(5)	Off by 2.5	80	-0.0372	-0.0368	-0.0367	-0.0368	-0.0371	-0.0368	0.0581	0.0573	0.0566	0.0561	0.0563	0.0562

Table 16 Bias and RMSE for 97.5th Percentile by Distribution, Test Targeting and Test Length for Sample Size 200

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
200	Normal	On-target	10	0.0179	0.0050	0.0079	0.0063	0.0062	0.0061	0.2935	0.2557	0.2490	0.2440	0.2414	0.2397
200	Normal	On-target	20	0.0049	0.0163	0.0173	0.0135	0.0132	0.0142	0.2424	0.2141	0.2020	0.1959	0.1941	0.1932
200	Normal	On-target	40	0.0206	0.0165	0.0176	0.0171	0.0173	0.0166	0.2175	0.1864	0.1775	0.1758	0.1734	0.1699
200	Normal	On-target	60	0.0134	0.0134	0.0127	0.0137	0.0140	0.0135	0.2148	0.1836	0.1780	0.1746	0.1722	0.1701
200	Normal	On-target	80	0.0106	0.0133	0.0098	0.0120	0.0116	0.0133	0.2115	0.1824	0.1791	0.1741	0.1707	0.1719
200	Normal	Off by 0.5	10	0.0055	-0.0047	-0.0033	-0.0018	-0.0020	-0.0026	0.2813	0.2472	0.2447	0.2337	0.2348	0.2316
200	Normal	Off by 0.5	20	0.0059	0.0086	0.0061	0.0039	0.0055	0.0071	0.2419	0.2059	0.1987	0.1917	0.1902	0.1896
200	Normal	Off by 0.5	40	0.0103	0.0122	0.0096	0.0082	0.0088	0.0096	0.2089	0.1826	0.1748	0.1732	0.1709	0.1692
200	Normal	Off by 0.5	60	0.0047	0.0061	0.0059	0.0067	0.0072	0.0061	0.2046	0.1786	0.1791	0.1743	0.1720	0.1703
200	Normal	Off by 0.5	80	0.0070	0.0056	0.0086	0.0078	0.0072	0.0082	0.2032	0.1772	0.1763	0.1707	0.1696	0.1714
200	Normal	Off by 1.0	10	-0.0125	-0.0123	-0.0109	-0.0128	-0.0146	-0.0142	0.2903	0.2516	0.2458	0.2416	0.2393	0.2382
200	Normal	Off by 1.0	20	-0.0101	0.0001	-0.0018	-0.0014	0.0027	0.0022	0.2347	0.2050	0.1949	0.1920	0.1904	0.1882
200	Normal	Off by 1.0	40	0.0049	-0.0027	0.0027	0.0029	0.0019	0.0041	0.2099	0.1796	0.1777	0.1728	0.1708	0.1692
200	Normal	Off by 1.0	60	-0.0058	-0.0023	-0.0034	-0.0006	-0.0033	-0.0004	0.2068	0.1825	0.1793	0.1768	0.1726	0.1724
200	Normal	Off by 1.0	80	-0.0038	-0.0018	-0.0019	-0.0015	-0.0016	-0.0015	0.2012	0.1790	0.1738	0.1691	0.1676	0.1677
200	Uniform	On-target	10	0.3502	0.3472	0.3447	0.3435	0.3444	0.3431	0.4467	0.4250	0.4200	0.4149	0.4154	0.4142
200	Uniform	On-target	20	0.3026	0.3065	0.3098	0.3103	0.3085	0.3066	0.3660	0.3498	0.3506	0.3461	0.3438	0.3416
200	Uniform	On-target	40	0.2359	0.2323	0.2331	0.2325	0.2317	0.2315	0.2821	0.2646	0.2622	0.2602	0.2569	0.2558
200	Uniform	On-target	60	0.1879	0.1933	0.1938	0.1921	0.1948	0.1939	0.2325	0.2245	0.2203	0.2170	0.2189	0.2177
200	Uniform	On-target	80	0.1618	0.1647	0.1627	0.1655	0.1653	0.1657	0.2050	0.1959	0.1905	0.1913	0.1899	0.1899
200	Uniform	Off by 0.5	10	0.3347	0.3328	0.3340	0.3309	0.3313	0.3324	0.4266	0.4073	0.4047	0.3960	0.3963	0.3965
200	Uniform	Off by 0.5	20	0.2980	0.2912	0.2922	0.2913	0.2907	0.2923	0.3582	0.3324	0.3294	0.3257	0.3235	0.3248
200	Uniform	Off by 0.5	40	0.2196	0.2174	0.2177	0.2184	0.2176	0.2168	0.2703	0.2506	0.2491	0.2453	0.2430	0.2423
200	Uniform	Off by 0.5	60	0.1741	0.1753	0.1782	0.1789	0.1797	0.1789	0.2203	0.2053	0.2062	0.2044	0.2041	0.2022
200	Uniform	Off by 0.5	80	0.1493	0.1502	0.1504	0.1523	0.1519	0.1507	0.1962	0.1801	0.1782	0.1779	0.1768	0.1752
200	Uniform	Off by 1.0	10	0.3178	0.3186	0.3234	0.3241	0.3202	0.3211	0.4109	0.3909	0.3889	0.3874	0.3835	0.3830
200	Uniform	Off by 1.0	20	0.2832	0.2774	0.2773	0.2756	0.2750	0.2762	0.3462	0.3196	0.3154	0.3104	0.3097	0.3094
200	Uniform	Off by 1.0	40	0.2067	0.2032	0.2029	0.2028	0.2018	0.2007	0.2571	0.2380	0.2336	0.2303	0.2286	0.2270
200	Uniform	Off by 1.0	60	0.1637	0.1634	0.1628	0.1645	0.1646	0.1631	0.2127	0.1954	0.1894	0.1886	0.1883	0.1862
200	Uniform	Off by 1.0	80	0.1353	0.1349	0.1353	0.1366	0.1373	0.1373	0.1801	0.1681	0.1643	0.1639	0.1634	0.1627
200	Bimodal	On-target	10	0.7521	0.7608	0.7547	0.7527	0.7554	0.7545	0.7935	0.7923	0.7833	0.7813	0.7829	0.7823
200	Bimodal	On-target	20	0.7122	0.7211	0.7196	0.7170	0.7166	0.7196	0.7385	0.7396	0.7358	0.7324	0.7318	0.7341
200	Bimodal	On-target	40	0.6328	0.6364	0.6371	0.6361	0.6359	0.6359	0.6521	0.6481	0.6478	0.6459	0.6454	0.6454
200	Bimodal	On-target	60	0.5911	0.5910	0.5974	0.5946	0.5935	0.5937	0.6072	0.6016	0.6068	0.6031	0.6018	0.6021
200	Bimodal	On-target	80	0.5610	0.5635	0.5640	0.5626	0.5623	0.5620	0.5769	0.5730	0.5728	0.5712	0.5702	0.5697
200	Bimodal	Off by 0.5	10	0.7364	0.7334	0.7393	0.7345	0.7407	0.7364	0.7762	0.7626	0.7667	0.7606	0.7663	0.7611
200	Bimodal	Off by 0.5	20	0.6974	0.7020	0.6987	0.6984	0.6964	0.6972	0.7226	0.7193	0.7138	0.7128	0.7104	0.7109
200	Bimodal	Off by 0.5	40	0.6242	0.6227	0.6201	0.6211	0.6207	0.6204	0.6424	0.6345	0.6308	0.6311	0.6301	0.6298
200	Bimodal	Off by 0.5	60	0.5817	0.5780	0.5790	0.5785	0.5774	0.5785	0.5978	0.5889	0.5887	0.5874	0.5858	0.5867
200	Bimodal	Off by 0.5	80	0.5474	0.5487	0.5514	0.5488	0.5491	0.5494	0.5625	0.5584	0.5602	0.5569	0.5569	0.5572
200	Bimodal	Off by 1.0	10	0.7275	0.7304	0.7231	0.7279	0.7294	0.7277	0.7667	0.7605	0.7503	0.7543	0.7544	0.7523
200	Bimodal	Off by 1.0	20	0.6860	0.6865	0.6857	0.6831	0.6824	0.6823	0.7111	0.7034	0.7020	0.6979	0.6966	0.6967

Appendix 3

Estimated Bias and RMSEs of Simulation

200	Bimodal	Off by 1.0	40	0.6031	0.6102	0.6080	0.6055	0.6055	0.6051	0.6213	0.6212	0.6183	0.6150	0.6145	0.6139
200	Bimodal	Off by 1.0	60	0.5686	0.5688	0.5689	0.5692	0.5667	0.5688	0.5847	0.5799	0.5789	0.5781	0.5753	0.5773
200	Bimodal	Off by 1.0	80	0.5380	0.5380	0.5368	0.5366	0.5370	0.5369	0.5534	0.5478	0.5459	0.5449	0.5452	0.5451
200	Chisq(5)	On-target	10	-0.4724	-0.4827	-0.4902	-0.4865	-0.4869	-0.4854	0.5778	0.5730	0.5760	0.5690	0.5683	0.5670
200	Chisq(5)	On-target	20	-0.3669	-0.3651	-0.3684	-0.3677	-0.3699	-0.3689	0.4843	0.4570	0.4572	0.4534	0.4533	0.4529
200	Chisq(5)	On-target	40	-0.2458	-0.2374	-0.2417	-0.2430	-0.2416	-0.2442	0.3912	0.3673	0.3674	0.3623	0.3633	0.3636
200	Chisq(5)	On-target	60	-0.1765	-0.1688	-0.1755	-0.1749	-0.1735	-0.1721	0.3598	0.3334	0.3324	0.3284	0.3270	0.3257
200	Chisq(5)	On-target	80	-0.1289	-0.1388	-0.1348	-0.1350	-0.1372	-0.1355	0.3347	0.3293	0.3199	0.3196	0.3181	0.3176
200	Chisq(5)	Off by 0.5	10	-0.4340	-0.4393	-0.4376	-0.4413	-0.4408	-0.4408	0.5561	0.5379	0.5344	0.5337	0.5324	0.5316
200	Chisq(5)	Off by 0.5	20	-0.3233	-0.3217	-0.3274	-0.3277	-0.3261	-0.3247	0.4466	0.4258	0.4251	0.4212	0.4189	0.4180
200	Chisq(5)	Off by 0.5	40	-0.2193	-0.2226	-0.2157	-0.2192	-0.2176	-0.2163	0.3893	0.3591	0.3573	0.3537	0.3530	0.3519
200	Chisq(5)	Off by 0.5	60	-0.1528	-0.1523	-0.1548	-0.1473	-0.1520	-0.1502	0.3623	0.3357	0.3322	0.3284	0.3259	0.3253
200	Chisq(5)	Off by 0.5	80	-0.1160	-0.1209	-0.1190	-0.1185	-0.1205	-0.1189	0.3436	0.3256	0.3218	0.3169	0.3193	0.3183
200	Chisq(5)	Off by 1.0	10	-0.4018	-0.3992	-0.4025	-0.4018	-0.4015	-0.4028	0.5267	0.5100	0.5056	0.5031	0.5015	0.5011
200	Chisq(5)	Off by 1.0	20	-0.3099	-0.2951	-0.2943	-0.2994	-0.3001	-0.3015	0.4422	0.4136	0.4093	0.4078	0.4072	0.4093
200	Chisq(5)	Off by 1.0	40	-0.1927	-0.1938	-0.1979	-0.1960	-0.1983	-0.1952	0.3807	0.3558	0.3530	0.3505	0.3490	0.3479
200	Chisq(5)	Off by 1.0	60	-0.1399	-0.1345	-0.1373	-0.1352	-0.1366	-0.1362	0.3649	0.3349	0.3315	0.3289	0.3267	0.3285
200	Chisq(5)	Off by 1.0	80	-0.1110	-0.1127	-0.1155	-0.1138	-0.1147	-0.1143	0.3506	0.3280	0.3247	0.3260	0.3253	0.3242

Table 17 Bias and RMSE for 97.5th Percentile by Distribution, Test Targeting and Test Length for Sample Size 2000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
2000	Normal	On-target	10	0.0024	0.0044	0.0033	0.0025	0.0025	0.0031	0.1009	0.0897	0.0877	0.0853	0.0852	0.0842
2000	Normal	On-target	20	-0.0022	0.0009	0.0002	-0.0001	0.0001	-0.0007	0.0842	0.0727	0.0701	0.0688	0.0682	0.0681
2000	Normal	On-target	40	0.0008	0.0015	0.0004	0.0008	-0.0007	0.0006	0.0754	0.0677	0.0655	0.0627	0.0630	0.0629
2000	Normal	On-target	60	0.0029	0.0025	0.0044	0.0038	0.0034	0.0037	0.0723	0.0627	0.0601	0.0595	0.0579	0.0584
2000	Normal	On-target	80	0.0016	0.0013	0.0015	0.0018	0.0007	0.0013	0.0705	0.0622	0.0597	0.0594	0.0582	0.0582
2000	Normal	Off by 0.5	10	-0.0022	-0.0008	-0.0006	-0.0012	-0.0019	-0.0012	0.0962	0.0844	0.0822	0.0812	0.0810	0.0800
2000	Normal	Off by 0.5	20	0.0001	-0.0030	-0.0028	-0.0031	-0.0025	-0.0031	0.0858	0.0728	0.0712	0.0686	0.0682	0.0674
2000	Normal	Off by 0.5	40	0.0019	-0.0029	-0.0004	0.0001	-0.0009	-0.0006	0.0721	0.0642	0.0615	0.0601	0.0598	0.0598
2000	Normal	Off by 0.5	60	-0.0024	-0.0027	-0.0019	-0.0027	-0.0016	-0.0025	0.0738	0.0644	0.0623	0.0606	0.0607	0.0612
2000	Normal	Off by 0.5	80	0.0052	0.0037	0.0033	0.0040	0.0033	0.0034	0.0712	0.0637	0.0617	0.0604	0.0601	0.0602
2000	Normal	Off by 1.0	10	-0.0002	-0.0025	-0.0008	-0.0007	-0.0014	-0.0015	0.1010	0.0880	0.0877	0.0844	0.0840	0.0834
2000	Normal	Off by 1.0	20	0.0009	-0.0032	-0.0028	-0.0033	-0.0036	-0.0032	0.0814	0.0716	0.0698	0.0682	0.0675	0.0672
2000	Normal	Off by 1.0	40	-0.0036	-0.0035	-0.0037	-0.0026	-0.0025	-0.0030	0.0750	0.0665	0.0635	0.0620	0.0624	0.0610
2000	Normal	Off by 1.0	60	0.0011	0.0009	0.0002	-0.0002	-0.0005	0.0001	0.0683	0.0615	0.0598	0.0587	0.0575	0.0576
2000	Normal	Off by 1.0	80	-0.0008	-0.0027	-0.0014	-0.0013	-0.0009	-0.0006	0.0689	0.0609	0.0596	0.0583	0.0577	0.0580
2000	Normal	Off by 2.5	10	-0.0063	-0.0079	-0.0072	-0.0073	-0.0086	-0.0074	0.1056	0.0975	0.0952	0.0920	0.0921	0.0915
2000	Normal	Off by 2.5	20	-0.0102	-0.0092	-0.0070	-0.0087	-0.0079	-0.0079	0.0882	0.0757	0.0736	0.0734	0.0724	0.0716
2000	Normal	Off by 2.5	40	-0.0069	-0.0062	-0.0076	-0.0085	-0.0075	-0.0085	0.0754	0.0666	0.0652	0.0633	0.0630	0.0634
2000	Normal	Off by 2.5	60	-0.0075	-0.0099	-0.0081	-0.0089	-0.0089	-0.0088	0.0741	0.0650	0.0628	0.0624	0.0617	0.0615
2000	Normal	Off by 2.5	80	-0.0100	-0.0099	-0.0088	-0.0098	-0.0100	-0.0096	0.0691	0.0630	0.0616	0.0598	0.0603	0.0598
2000	Uniform	On-target	10	0.3220	0.3239	0.3236	0.3250	0.3236	0.3242	0.3354	0.3339	0.3330	0.3341	0.3327	0.3331
2000	Uniform	On-target	20	0.2885	0.2892	0.2884	0.2884	0.2886	0.2885	0.2982	0.2958	0.2946	0.2944	0.2943	0.2942
2000	Uniform	On-target	40	0.2206	0.2218	0.2215	0.2230	0.2226	0.2231	0.2282	0.2274	0.2267	0.2278	0.2273	0.2277
2000	Uniform	On-target	60	0.1845	0.1825	0.1829	0.1831	0.1829	0.1828	0.1916	0.1878	0.1877	0.1877	0.1874	0.1874
2000	Uniform	On-target	80	0.1586	0.1575	0.1575	0.1574	0.1573	0.1580	0.1656	0.1625	0.1621	0.1617	0.1616	0.1622
2000	Uniform	Off by 0.5	10	0.3225	0.3236	0.3233	0.3231	0.3242	0.3235	0.3353	0.3340	0.3332	0.3325	0.3334	0.3326
2000	Uniform	Off by 0.5	20	0.2768	0.2756	0.2767	0.2765	0.2769	0.2770	0.2863	0.2825	0.2829	0.2825	0.2828	0.2828
2000	Uniform	Off by 0.5	40	0.2103	0.2110	0.2115	0.2104	0.2108	0.2105	0.2181	0.2164	0.2164	0.2150	0.2155	0.2149
2000	Uniform	Off by 0.5	60	0.1749	0.1728	0.1741	0.1740	0.1739	0.1736	0.1823	0.1778	0.1788	0.1784	0.1783	0.1779
2000	Uniform	Off by 0.5	80	0.1470	0.1452	0.1449	0.1450	0.1458	0.1454	0.1542	0.1504	0.1497	0.1498	0.1502	0.1498
2000	Uniform	Off by 1.0	10	0.3168	0.3169	0.3203	0.3196	0.3185	0.3184	0.3293	0.3264	0.3294	0.3280	0.3269	0.3268
2000	Uniform	Off by 1.0	20	0.2737	0.2738	0.2729	0.2737	0.2727	0.2733	0.2836	0.2802	0.2790	0.2798	0.2785	0.2790
2000	Uniform	Off by 1.0	40	0.2055	0.2071	0.2066	0.2063	0.2069	0.2065	0.2129	0.2128	0.2116	0.2110	0.2117	0.2111
2000	Uniform	Off by 1.0	60	0.1669	0.1672	0.1664	0.1668	0.1665	0.1664	0.1745	0.1722	0.1713	0.1712	0.1708	0.1708
2000	Uniform	Off by 1.0	80	0.1414	0.1396	0.1395	0.1393	0.1399	0.1395	0.1488	0.1448	0.1444	0.1439	0.1446	0.1440
2000	Uniform	Off by 2.5	10	0.2899	0.2910	0.2906	0.2913	0.2914	0.2907	0.3045	0.3029	0.3022	0.3023	0.3024	0.3014
2000	Uniform	Off by 2.5	20	0.2571	0.2558	0.2562	0.2560	0.2560	0.2558	0.2669	0.2639	0.2636	0.2629	0.2629	0.2626
2000	Uniform	Off by 2.5	40	0.1974	0.1963	0.1969	0.1966	0.1977	0.1970	0.2053	0.2022	0.2023	0.2015	0.2025	0.2018
2000	Uniform	Off by 2.5	60	0.1590	0.1593	0.1595	0.1594	0.1594	0.1590	0.1668	0.1650	0.1647	0.1643	0.1643	0.1639
2000	Uniform	Off by 2.5	80	0.1289	0.1298	0.1310	0.1298	0.1299	0.1297	0.1367	0.1354	0.1360	0.1347	0.1347	0.1345
2000	Bimodal	On-target	10	0.7535	0.7499	0.7507	0.7505	0.7514	0.7508	0.7590	0.7541	0.7546	0.7543	0.7551	0.7544
2000	Bimodal	On-target	20	0.7043	0.7068	0.7064	0.7066	0.7057	0.7058	0.7079	0.7095	0.7088	0.7090	0.7080	0.7081

Appendix 3

Estimated Bias and RMSEs of Simulation

2000	Bimodal	On-target	40	0.6293	0.6272	0.6278	0.6280	0.6281	0.6284	0.6320	0.6292	0.6295	0.6296	0.6297	0.6299
2000	Bimodal	On-target	60	0.5809	0.5822	0.5822	0.5814	0.5818	0.5816	0.5832	0.5839	0.5837	0.5828	0.5832	0.5829
2000	Bimodal	On-target	80	0.5511	0.5523	0.5518	0.5521	0.5524	0.5522	0.5532	0.5537	0.5532	0.5534	0.5536	0.5535
2000	Bimodal	Off by 0.5	10	0.7439	0.7425	0.7440	0.7425	0.7426	0.7425	0.7493	0.7464	0.7476	0.7459	0.7460	0.7459
2000	Bimodal	Off by 0.5	20	0.6969	0.6970	0.6963	0.6959	0.6957	0.6962	0.7006	0.6996	0.6987	0.6981	0.6978	0.6983
2000	Bimodal	Off by 0.5	40	0.6186	0.6187	0.6200	0.6184	0.6190	0.6186	0.6212	0.6205	0.6217	0.6200	0.6205	0.6201
2000	Bimodal	Off by 0.5	60	0.5735	0.5736	0.5733	0.5731	0.5730	0.5730	0.5757	0.5752	0.5747	0.5744	0.5743	0.5742
2000	Bimodal	Off by 0.5	80	0.5462	0.5453	0.5447	0.5443	0.5448	0.5449	0.5482	0.5467	0.5460	0.5455	0.5460	0.5461
2000	Bimodal	Off by 1.0	10	0.7395	0.7411	0.7412	0.7410	0.7403	0.7407	0.7448	0.7452	0.7450	0.7447	0.7438	0.7441
2000	Bimodal	Off by 1.0	20	0.6904	0.6891	0.6874	0.6872	0.6875	0.6878	0.6939	0.6917	0.6897	0.6893	0.6896	0.6899
2000	Bimodal	Off by 1.0	40	0.6100	0.6104	0.6121	0.6109	0.6103	0.6104	0.6126	0.6120	0.6137	0.6124	0.6117	0.6119
2000	Bimodal	Off by 1.0	60	0.5663	0.5645	0.5652	0.5657	0.5656	0.5655	0.5684	0.5660	0.5666	0.5670	0.5668	0.5667
2000	Bimodal	Off by 1.0	80	0.5402	0.5391	0.5389	0.5387	0.5389	0.5389	0.5422	0.5405	0.5402	0.5400	0.5401	0.5401
2000	Bimodal	Off by 2.5	10	0.7202	0.7181	0.7173	0.7186	0.7184	0.7183	0.7261	0.7230	0.7218	0.7230	0.7228	0.7226
2000	Bimodal	Off by 2.5	20	0.6748	0.6747	0.6738	0.6735	0.6735	0.6740	0.6787	0.6776	0.6764	0.6760	0.6759	0.6764
2000	Bimodal	Off by 2.5	40	0.6092	0.6085	0.6088	0.6080	0.6079	0.6078	0.6120	0.6104	0.6105	0.6097	0.6096	0.6095
2000	Bimodal	Off by 2.5	60	0.5625	0.5608	0.5605	0.5617	0.5612	0.5618	0.5648	0.5624	0.5621	0.5632	0.5627	0.5632
2000	Bimodal	Off by 2.5	80	0.5332	0.5328	0.5327	0.5334	0.5335	0.5332	0.5354	0.5343	0.5341	0.5347	0.5348	0.5345
2000	Chisq(5)	On-target	10	-0.4914	-0.4916	-0.4917	-0.4926	-0.4920	-0.4923	0.5052	0.5035	0.5030	0.5037	0.5030	0.5033
2000	Chisq(5)	On-target	20	-0.3837	-0.3857	-0.3845	-0.3848	-0.3850	-0.3859	0.4006	0.3996	0.3977	0.3975	0.3974	0.3982
2000	Chisq(5)	On-target	40	-0.2464	-0.2447	-0.2439	-0.2443	-0.2445	-0.2436	0.2722	0.2662	0.2646	0.2645	0.2646	0.2634
2000	Chisq(5)	On-target	60	-0.1790	-0.1799	-0.1782	-0.1787	-0.1789	-0.1790	0.2122	0.2080	0.2053	0.2050	0.2050	0.2050
2000	Chisq(5)	On-target	80	-0.1294	-0.1305	-0.1288	-0.1306	-0.1309	-0.1300	0.1718	0.1667	0.1644	0.1653	0.1654	0.1641
2000	Chisq(5)	Off by 0.5	10	-0.4472	-0.4441	-0.4429	-0.4452	-0.4439	-0.4435	0.4611	0.4566	0.4550	0.4567	0.4554	0.4549
2000	Chisq(5)	Off by 0.5	20	-0.3346	-0.3307	-0.3331	-0.3327	-0.3320	-0.3329	0.3539	0.3467	0.3491	0.3477	0.3469	0.3477
2000	Chisq(5)	Off by 0.5	40	-0.2122	-0.2151	-0.2134	-0.2138	-0.2129	-0.2139	0.2432	0.2411	0.2375	0.2372	0.2367	0.2373
2000	Chisq(5)	Off by 0.5	60	-0.1496	-0.1524	-0.1520	-0.1528	-0.1529	-0.1528	0.1895	0.1863	0.1852	0.1845	0.1842	0.1846
2000	Chisq(5)	Off by 0.5	80	-0.1143	-0.1154	-0.1152	-0.1159	-0.1158	-0.1155	0.1612	0.1537	0.1537	0.1523	0.1524	0.1522
2000	Chisq(5)	Off by 1.0	10	-0.3942	-0.3945	-0.3940	-0.3947	-0.3945	-0.3948	0.4120	0.4087	0.4076	0.4078	0.4078	0.4079
2000	Chisq(5)	Off by 1.0	20	-0.2850	-0.2908	-0.2907	-0.2901	-0.2899	-0.2896	0.3078	0.3103	0.3087	0.3080	0.3076	0.3071
2000	Chisq(5)	Off by 1.0	40	-0.1848	-0.1865	-0.1843	-0.1836	-0.1846	-0.1848	0.2190	0.2145	0.2124	0.2107	0.2108	0.2111
2000	Chisq(5)	Off by 1.0	60	-0.1333	-0.1332	-0.1323	-0.1319	-0.1320	-0.1322	0.1804	0.1719	0.1709	0.1696	0.1693	0.1692
2000	Chisq(5)	Off by 1.0	80	-0.1004	-0.0981	-0.0980	-0.0979	-0.0980	-0.0991	0.1566	0.1481	0.1481	0.1463	0.1463	0.1467
2000	Chisq(5)	Off by 2.5	10	-0.2480	-0.2464	-0.2480	-0.2475	-0.2484	-0.2486	0.2791	0.2728	0.2729	0.2720	0.2723	0.2725
2000	Chisq(5)	Off by 2.5	20	-0.2014	-0.1973	-0.1980	-0.1986	-0.1978	-0.1975	0.2349	0.2253	0.2246	0.2242	0.2233	0.2228
2000	Chisq(5)	Off by 2.5	40	-0.1303	-0.1334	-0.1320	-0.1318	-0.1324	-0.1323	0.1800	0.1755	0.1727	0.1714	0.1723	0.1723
2000	Chisq(5)	Off by 2.5	60	-0.0986	-0.1016	-0.1030	-0.1016	-0.1017	-0.1017	0.1534	0.1502	0.1498	0.1479	0.1480	0.1473
2000	Chisq(5)	Off by 2.5	80	-0.0784	-0.0798	-0.0777	-0.0793	-0.0795	-0.0794	0.1390	0.1356	0.1333	0.1327	0.1324	0.1329

Table 18 Bias and RMSE for 97.5th Percentile by Distribution, Test Targeting and Test Length for Sample Size 10,000

Sample Size	Distribution	Test Target	Test Length	Bias						RMSE					
				1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs	1 PV	3 PVs	5 PVs	10 PVs	15 PVs	20 PVs
10000	Normal	On-target	10	0.0001	0.0007	0.0006	0.0009	0.0002	0.0006	0.0598	0.0563	0.0551	0.0549	0.0542	0.0542
10000	Normal	On-target	20	-0.0009	0.0003	-0.0001	0.0003	0.0005	0.0001	0.0520	0.0481	0.0470	0.0469	0.0467	0.0463
10000	Normal	On-target	40	-0.0001	0.0006	0.0005	0.0009	0.0009	0.0010	0.0469	0.0432	0.0424	0.0423	0.0419	0.0419
10000	Normal	On-target	60	0.0021	0.0019	0.0023	0.0021	0.0021	0.0022	0.0451	0.0406	0.0401	0.0397	0.0394	0.0393
10000	Normal	On-target	80	0.0035	0.0032	0.0035	0.0031	0.0031	0.0031	0.0411	0.0390	0.0380	0.0375	0.0373	0.0373
10000	Normal	Off by 0.5	10	-0.0008	-0.0007	0.0001	0.0000	0.0001	0.0000	0.0578	0.0531	0.0530	0.0523	0.0521	0.0517
10000	Normal	Off by 0.5	20	0.0000	0.0005	0.0006	0.0006	0.0007	0.0009	0.0503	0.0478	0.0466	0.0464	0.0459	0.0459
10000	Normal	Off by 0.5	40	0.0002	0.0009	0.0011	0.0010	0.0008	0.0010	0.0472	0.0433	0.0423	0.0416	0.0412	0.0414
10000	Normal	Off by 0.5	60	0.0034	0.0025	0.0020	0.0023	0.0027	0.0022	0.0418	0.0396	0.0390	0.0389	0.0388	0.0387
10000	Normal	Off by 0.5	80	0.0025	0.0037	0.0040	0.0034	0.0035	0.0036	0.0423	0.0386	0.0384	0.0380	0.0380	0.0376
10000	Normal	Off by 1.0	10	0.0000	0.0000	-0.0005	-0.0002	-0.0005	-0.0005	0.0585	0.0539	0.0542	0.0534	0.0530	0.0529
10000	Normal	Off by 1.0	20	0.0007	0.0007	0.0010	0.0013	0.0010	0.0012	0.0532	0.0475	0.0468	0.0464	0.0460	0.0461
10000	Normal	Off by 1.0	40	0.0007	0.0008	0.0002	0.0011	0.0007	0.0005	0.0457	0.0425	0.0412	0.0411	0.0407	0.0407
10000	Normal	Off by 1.0	60	0.0027	0.0019	0.0015	0.0014	0.0018	0.0017	0.0420	0.0395	0.0384	0.0380	0.0380	0.0378
10000	Normal	Off by 1.0	80	0.0013	0.0023	0.0021	0.0023	0.0024	0.0023	0.0427	0.0384	0.0381	0.0374	0.0375	0.0375
10000	Normal	Off by 2.5	10	-0.0007	-0.0019	-0.0014	-0.0016	-0.0017	-0.0014	0.0584	0.0544	0.0548	0.0539	0.0534	0.0536
10000	Normal	Off by 2.5	20	0.0019	-0.0008	-0.0013	-0.0005	-0.0003	-0.0009	0.0529	0.0485	0.0475	0.0472	0.0475	0.0468
10000	Normal	Off by 2.5	40	-0.0018	-0.0013	-0.0005	-0.0005	-0.0005	-0.0010	0.0474	0.0426	0.0419	0.0417	0.0415	0.0413
10000	Normal	Off by 2.5	60	-0.0009	-0.0018	-0.0016	-0.0014	-0.0013	-0.0013	0.0425	0.0393	0.0395	0.0390	0.0388	0.0389
10000	Normal	Off by 2.5	80	-0.0004	0.0001	0.0003	0.0002	0.0004	0.0003	0.0423	0.0381	0.0373	0.0369	0.0370	0.0370
10000	Uniform	On-target	10	0.3298	0.3289	0.3283	0.3287	0.3283	0.3284	0.3349	0.3332	0.3326	0.3329	0.3325	0.3325
10000	Uniform	On-target	20	0.2901	0.2890	0.2884	0.2897	0.2889	0.2894	0.2938	0.2924	0.2915	0.2928	0.2919	0.2924
10000	Uniform	On-target	40	0.2227	0.2234	0.2237	0.2234	0.2236	0.2235	0.2261	0.2260	0.2262	0.2259	0.2261	0.2260
10000	Uniform	On-target	60	0.1829	0.1826	0.1828	0.1827	0.1826	0.1826	0.1861	0.1852	0.1853	0.1851	0.1850	0.1850
10000	Uniform	On-target	80	0.1554	0.1552	0.1555	0.1553	0.1555	0.1552	0.1585	0.1578	0.1579	0.1577	0.1579	0.1575
10000	Uniform	Off by 0.5	10	0.3214	0.3226	0.3225	0.3229	0.3227	0.3226	0.3263	0.3270	0.3268	0.3269	0.3269	0.3267
10000	Uniform	Off by 0.5	20	0.2806	0.2794	0.2795	0.2799	0.2796	0.2796	0.2843	0.2825	0.2827	0.2828	0.2826	0.2825
10000	Uniform	Off by 0.5	40	0.2124	0.2122	0.2124	0.2121	0.2126	0.2126	0.2158	0.2150	0.2150	0.2147	0.2152	0.2152
10000	Uniform	Off by 0.5	60	0.1706	0.1711	0.1716	0.1715	0.1714	0.1715	0.1739	0.1738	0.1742	0.1740	0.1738	0.1740
10000	Uniform	Off by 0.5	80	0.1447	0.1452	0.1450	0.1450	0.1451	0.1449	0.1477	0.1477	0.1475	0.1474	0.1475	0.1473
10000	Uniform	Off by 1.0	10	0.3190	0.3194	0.3184	0.3190	0.3187	0.3186	0.3238	0.3235	0.3224	0.3230	0.3226	0.3224
10000	Uniform	Off by 1.0	20	0.2733	0.2739	0.2739	0.2739	0.2738	0.2742	0.2771	0.2771	0.2770	0.2770	0.2769	0.2772
10000	Uniform	Off by 1.0	40	0.2056	0.2061	0.2062	0.2063	0.2065	0.2066	0.2087	0.2087	0.2087	0.2086	0.2089	0.2089
10000	Uniform	Off by 1.0	60	0.1648	0.1651	0.1657	0.1653	0.1654	0.1652	0.1680	0.1677	0.1683	0.1677	0.1679	0.1676
10000	Uniform	Off by 1.0	80	0.1388	0.1386	0.1388	0.1384	0.1385	0.1384	0.1419	0.1412	0.1413	0.1409	0.1409	0.1408
10000	Uniform	Off by 2.5	10	0.3017	0.3002	0.3001	0.3004	0.3006	0.3004	0.3070	0.3049	0.3045	0.3048	0.3049	0.3047
10000	Uniform	Off by 2.5	20	0.2652	0.2659	0.2652	0.2657	0.2656	0.2657	0.2692	0.2694	0.2687	0.2689	0.2688	0.2689
10000	Uniform	Off by 2.5	40	0.2024	0.2036	0.2038	0.2031	0.2032	0.2034	0.2057	0.2062	0.2063	0.2055	0.2056	0.2059
10000	Uniform	Off by 2.5	60	0.1629	0.1632	0.1635	0.1630	0.1629	0.1630	0.1661	0.1658	0.1660	0.1655	0.1653	0.1654
10000	Uniform	Off by 2.5	80	0.1355	0.1357	0.1355	0.1355	0.1353	0.1353	0.1387	0.1385	0.1382	0.1381	0.1380	0.1379
10000	Bimodal	On-target	10	0.7534	0.7538	0.7538	0.7537	0.7538	0.7538	0.7556	0.7557	0.7556	0.7555	0.7556	0.7555
10000	Bimodal	On-target	20	0.7074	0.7070	0.7070	0.7068	0.7068	0.7070	0.7090	0.7084	0.7083	0.7080	0.7081	0.7083

Appendix 3

Estimated Bias and RMSEs of Simulation

10000	Bimodal	On-target	40	0.6304	0.6299	0.6298	0.6298	0.6299	0.6299	0.6315	0.6308	0.6307	0.6306	0.6308	0.6307
10000	Bimodal	On-target	60	0.5822	0.5826	0.5822	0.5823	0.5824	0.5824	0.5832	0.5834	0.5829	0.5831	0.5832	0.5831
10000	Bimodal	On-target	80	0.5523	0.5521	0.5524	0.5523	0.5524	0.5522	0.5531	0.5527	0.5531	0.5530	0.5530	0.5529
10000	Bimodal	Off by 0.5	10	0.7480	0.7458	0.7470	0.7460	0.7465	0.7466	0.7502	0.7476	0.7488	0.7478	0.7483	0.7483
10000	Bimodal	Off by 0.5	20	0.6962	0.6970	0.6975	0.6969	0.6971	0.6969	0.6977	0.6983	0.6988	0.6981	0.6983	0.6981
10000	Bimodal	Off by 0.5	40	0.6183	0.6189	0.6188	0.6185	0.6187	0.6183	0.6194	0.6198	0.6196	0.6193	0.6195	0.6191
10000	Bimodal	Off by 0.5	60	0.5722	0.5721	0.5722	0.5722	0.5723	0.5724	0.5731	0.5728	0.5729	0.5729	0.5730	0.5731
10000	Bimodal	Off by 0.5	80	0.5432	0.5434	0.5434	0.5432	0.5433	0.5435	0.5440	0.5440	0.5440	0.5439	0.5440	0.5441
10000	Bimodal	Off by 1.0	10	0.7443	0.7425	0.7430	0.7426	0.7430	0.7426	0.7464	0.7443	0.7448	0.7443	0.7447	0.7442
10000	Bimodal	Off by 1.0	20	0.6915	0.6913	0.6917	0.6919	0.6920	0.6918	0.6930	0.6925	0.6929	0.6930	0.6932	0.6929
10000	Bimodal	Off by 1.0	40	0.6124	0.6122	0.6130	0.6126	0.6125	0.6124	0.6135	0.6130	0.6138	0.6134	0.6133	0.6132
10000	Bimodal	Off by 1.0	60	0.5668	0.5664	0.5666	0.5669	0.5668	0.5667	0.5677	0.5672	0.5673	0.5676	0.5675	0.5674
10000	Bimodal	Off by 1.0	80	0.5375	0.5382	0.5386	0.5382	0.5381	0.5382	0.5383	0.5389	0.5393	0.5388	0.5387	0.5388
10000	Bimodal	Off by 2.5	10	0.7273	0.7259	0.7266	0.7261	0.7265	0.7264	0.7293	0.7278	0.7284	0.7278	0.7283	0.7282
10000	Bimodal	Off by 2.5	20	0.6847	0.6839	0.6840	0.6847	0.6842	0.6842	0.6863	0.6852	0.6853	0.6859	0.6855	0.6855
10000	Bimodal	Off by 2.5	40	0.6135	0.6128	0.6122	0.6129	0.6122	0.6124	0.6146	0.6136	0.6130	0.6138	0.6131	0.6133
10000	Bimodal	Off by 2.5	60	0.5674	0.5671	0.5667	0.5672	0.5669	0.5669	0.5683	0.5678	0.5674	0.5679	0.5676	0.5676
10000	Bimodal	Off by 2.5	80	0.5378	0.5377	0.5375	0.5377	0.5376	0.5374	0.5387	0.5384	0.5382	0.5383	0.5382	0.5381
10000	Chisq(5)	On-target	10	-0.4916	-0.4927	-0.4919	-0.4920	-0.4927	-0.4918	0.4963	0.4967	0.4959	0.4960	0.4967	0.4958
10000	Chisq(5)	On-target	20	-0.3718	-0.3722	-0.3717	-0.3724	-0.3721	-0.3721	0.3792	0.3783	0.3777	0.3782	0.3779	0.3779
10000	Chisq(5)	On-target	40	-0.2414	-0.2421	-0.2423	-0.2430	-0.2427	-0.2434	0.2512	0.2513	0.2510	0.2515	0.2513	0.2518
10000	Chisq(5)	On-target	60	-0.1774	-0.1761	-0.1755	-0.1760	-0.1759	-0.1762	0.1907	0.1880	0.1873	0.1877	0.1873	0.1875
10000	Chisq(5)	On-target	80	-0.1361	-0.1349	-0.1356	-0.1346	-0.1350	-0.1351	0.1530	0.1503	0.1502	0.1491	0.1495	0.1497
10000	Chisq(5)	Off by 0.5	10	-0.4366	-0.4389	-0.4382	-0.4384	-0.4385	-0.4379	0.4417	0.4436	0.4428	0.4430	0.4430	0.4424
10000	Chisq(5)	Off by 0.5	20	-0.3307	-0.3293	-0.3301	-0.3297	-0.3301	-0.3297	0.3385	0.3360	0.3367	0.3360	0.3365	0.3359
10000	Chisq(5)	Off by 0.5	40	-0.2084	-0.2079	-0.2080	-0.2085	-0.2082	-0.2084	0.2197	0.2183	0.2181	0.2183	0.2181	0.2182
10000	Chisq(5)	Off by 0.5	60	-0.1504	-0.1508	-0.1489	-0.1498	-0.1500	-0.1495	0.1657	0.1643	0.1623	0.1627	0.1631	0.1624
10000	Chisq(5)	Off by 0.5	80	-0.1165	-0.1154	-0.1155	-0.1151	-0.1152	-0.1150	0.1351	0.1322	0.1322	0.1317	0.1319	0.1315
10000	Chisq(5)	Off by 1.0	10	-0.3903	-0.3904	-0.3900	-0.3895	-0.3900	-0.3903	0.3962	0.3957	0.3953	0.3946	0.3951	0.3954
10000	Chisq(5)	Off by 1.0	20	-0.2895	-0.2903	-0.2898	-0.2900	-0.2900	-0.2901	0.2983	0.2978	0.2972	0.2971	0.2971	0.2971
10000	Chisq(5)	Off by 1.0	40	-0.1825	-0.1823	-0.1821	-0.1828	-0.1827	-0.1825	0.1952	0.1941	0.1934	0.1938	0.1937	0.1935
10000	Chisq(5)	Off by 1.0	60	-0.1342	-0.1316	-0.1315	-0.1322	-0.1321	-0.1318	0.1508	0.1470	0.1465	0.1467	0.1465	0.1463
10000	Chisq(5)	Off by 1.0	80	-0.1030	-0.1023	-0.1019	-0.1021	-0.1020	-0.1020	0.1237	0.1211	0.1201	0.1201	0.1198	0.1199
10000	Chisq(5)	Off by 2.5	10	-0.2381	-0.2382	-0.2377	-0.2377	-0.2380	-0.2381	0.2487	0.2479	0.2470	0.2467	0.2469	0.2470
10000	Chisq(5)	Off by 2.5	20	-0.1865	-0.1878	-0.1879	-0.1877	-0.1880	-0.1876	0.2001	0.1993	0.1990	0.1985	0.1989	0.1983
10000	Chisq(5)	Off by 2.5	40	-0.1259	-0.1265	-0.1257	-0.1268	-0.1264	-0.1265	0.1432	0.1418	0.1408	0.1412	0.1409	0.1410
10000	Chisq(5)	Off by 2.5	60	-0.0968	-0.0957	-0.0958	-0.0964	-0.0959	-0.0961	0.1186	0.1159	0.1162	0.1161	0.1156	0.1156
10000	Chisq(5)	Off by 2.5	80	-0.0749	-0.0749	-0.0749	-0.0751	-0.0753	-0.0753	0.1009	0.0982	0.0972	0.0975	0.0974	0.0974

Appendix 4 Standard Error Estimates of Simulation One

Table 19 Standard Error Estimated for Mean by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.0888	0.0827	0.0782	0.0760	0.0750	0.0281	0.0262	0.0244	0.0240	0.0235	0.0125	0.0117	0.0109	0.0107	0.0106
Normal	On-target	5	0.0879	0.0820	0.0773	0.0757	0.0744	0.0277	0.0258	0.0243	0.0237	0.0234	0.0123	0.0115	0.0109	0.0106	0.0105
Normal	On-target	10	0.0864	0.0809	0.0770	0.0752	0.0742	0.0271	0.0254	0.0241	0.0237	0.0234	0.0122	0.0114	0.0108	0.0106	0.0104
Normal	On-target	15	0.0859	0.0806	0.0768	0.0751	0.0742	0.0271	0.0253	0.0241	0.0236	0.0233	0.0120	0.0113	0.0108	0.0106	0.0104
Normal	On-target	20	0.0856	0.0804	0.0767	0.0750	0.0741	0.0270	0.0253	0.0240	0.0236	0.0233	0.0120	0.0113	0.0108	0.0105	0.0104
Normal	Off by 0.5	3	0.0895	0.0834	0.0785	0.0763	0.0748	0.0279	0.0261	0.0245	0.0239	0.0236	0.0125	0.0116	0.0110	0.0107	0.0106
Normal	Off by 0.5	5	0.0874	0.0813	0.0774	0.0756	0.0745	0.0274	0.0257	0.0243	0.0237	0.0235	0.0123	0.0115	0.0109	0.0106	0.0105
Normal	Off by 0.5	10	0.0861	0.0812	0.0769	0.0752	0.0742	0.0272	0.0255	0.0241	0.0237	0.0234	0.0122	0.0114	0.0108	0.0106	0.0105
Normal	Off by 0.5	15	0.0859	0.0807	0.0767	0.0751	0.0742	0.0270	0.0254	0.0241	0.0236	0.0234	0.0121	0.0113	0.0108	0.0106	0.0104
Normal	Off by 0.5	20	0.0858	0.0806	0.0767	0.0751	0.0742	0.0269	0.0253	0.0241	0.0236	0.0234	0.0121	0.0113	0.0108	0.0106	0.0104
Normal	Off by 1.0	3	0.0895	0.0839	0.0782	0.0761	0.0750	0.0282	0.0261	0.0246	0.0240	0.0236	0.0126	0.0118	0.0110	0.0107	0.0106
Normal	Off by 1.0	5	0.0876	0.0826	0.0776	0.0758	0.0747	0.0277	0.0257	0.0244	0.0238	0.0235	0.0123	0.0115	0.0109	0.0106	0.0105
Normal	Off by 1.0	10	0.0865	0.0815	0.0770	0.0753	0.0742	0.0274	0.0255	0.0243	0.0237	0.0234	0.0121	0.0114	0.0108	0.0106	0.0105
Normal	Off by 1.0	15	0.0860	0.0812	0.0769	0.0753	0.0743	0.0271	0.0254	0.0242	0.0237	0.0234	0.0121	0.0114	0.0108	0.0106	0.0105
Normal	Off by 1.0	20	0.0854	0.0811	0.0769	0.0752	0.0742	0.0271	0.0254	0.0241	0.0237	0.0234	0.0121	0.0114	0.0108	0.0106	0.0104
Normal	Off by 2.5	3						0.0290	0.0272	0.0253	0.0246	0.0242	0.0129	0.0120	0.0114	0.0109	0.0108
Normal	Off by 2.5	5						0.0286	0.0266	0.0250	0.0244	0.0240	0.0127	0.0119	0.0112	0.0109	0.0107
Normal	Off by 2.5	10						0.0279	0.0263	0.0249	0.0242	0.0238	0.0125	0.0118	0.0111	0.0108	0.0107
Normal	Off by 2.5	15						0.0278	0.0262	0.0248	0.0242	0.0238	0.0124	0.0117	0.0111	0.0108	0.0106
Normal	Off by 2.5	20						0.0278	0.0262	0.0248	0.0241	0.0237	0.0124	0.0117	0.0111	0.0108	0.0106
Uniform	On-target	3	0.0911	0.0839	0.0790	0.0765	0.0757	0.0284	0.0263	0.0247	0.0240	0.0237	0.0126	0.0117	0.0111	0.0107	0.0106
Uniform	On-target	5	0.0888	0.0830	0.0778	0.0761	0.0753	0.0278	0.0258	0.0244	0.0239	0.0236	0.0125	0.0116	0.0109	0.0107	0.0105
Uniform	On-target	10	0.0872	0.0822	0.0775	0.0758	0.0750	0.0274	0.0256	0.0243	0.0238	0.0235	0.0123	0.0115	0.0109	0.0106	0.0105
Uniform	On-target	15	0.0872	0.0815	0.0772	0.0757	0.0748	0.0272	0.0256	0.0243	0.0237	0.0235	0.0122	0.0114	0.0108	0.0106	0.0105
Uniform	On-target	20	0.0869	0.0815	0.0773	0.0756	0.0748	0.0272	0.0255	0.0242	0.0237	0.0235	0.0121	0.0114	0.0108	0.0106	0.0105
Uniform	Off by 0.5	3	0.0908	0.0847	0.0788	0.0766	0.0757	0.0283	0.0262	0.0247	0.0241	0.0237	0.0127	0.0118	0.0111	0.0107	0.0106
Uniform	Off by 0.5	5	0.0890	0.0834	0.0783	0.0763	0.0753	0.0279	0.0260	0.0245	0.0240	0.0236	0.0125	0.0116	0.0110	0.0107	0.0105
Uniform	Off by 0.5	10	0.0875	0.0821	0.0777	0.0760	0.0751	0.0276	0.0257	0.0244	0.0238	0.0235	0.0123	0.0115	0.0109	0.0106	0.0105
Uniform	Off by 0.5	15	0.0875	0.0820	0.0776	0.0758	0.0749	0.0275	0.0256	0.0243	0.0238	0.0235	0.0122	0.0115	0.0109	0.0106	0.0105
Uniform	Off by 0.5	20	0.0873	0.0819	0.0774	0.0758	0.0749	0.0273	0.0256	0.0243	0.0237	0.0235	0.0122	0.0114	0.0109	0.0106	0.0105
Uniform	Off by 1.0	3	0.0921	0.0849	0.0793	0.0769	0.0760	0.0285	0.0265	0.0249	0.0242	0.0239	0.0128	0.0118	0.0111	0.0108	0.0106
Uniform	Off by 1.0	5	0.0893	0.0837	0.0786	0.0768	0.0755	0.0282	0.0262	0.0246	0.0241	0.0237	0.0126	0.0117	0.0110	0.0107	0.0106
Uniform	Off by 1.0	10	0.0882	0.0829	0.0780	0.0762	0.0752	0.0276	0.0259	0.0246	0.0239	0.0236	0.0123	0.0116	0.0109	0.0107	0.0105
Uniform	Off by 1.0	15	0.0882	0.0825	0.0779	0.0761	0.0751	0.0275	0.0258	0.0245	0.0239	0.0236	0.0123	0.0115	0.0109	0.0107	0.0105
Uniform	Off by 1.0	20	0.0877	0.0823	0.0778	0.0761	0.0751	0.0274	0.0258	0.0245	0.0239	0.0236	0.0122	0.0115	0.0109	0.0107	0.0105
Uniform	Off by 2.5	3						0.0295	0.0275	0.0257	0.0250	0.0247	0.0130	0.0124	0.0116	0.0112	0.0110
Uniform	Off by 2.5	5						0.0285	0.0270	0.0256	0.0248	0.0245	0.0129	0.0121	0.0115	0.0111	0.0109

Appendix 4

Standard Error Estimates of Simulation

Uniform	Off by 2.5	10						0.0282	0.0268	0.0254	0.0248	0.0243	0.0126	0.0119	0.0113	0.0110	0.0109
Uniform	Off by 2.5	15						0.0281	0.0267	0.0253	0.0247	0.0243	0.0125	0.0119	0.0113	0.0110	0.0109
Uniform	Off by 2.5	20						0.0279	0.0265	0.0252	0.0246	0.0243	0.0125	0.0119	0.0113	0.0110	0.0108
Bimodal	On-target	3	0.0902	0.0834	0.0786	0.0764	0.0754	0.0280	0.0263	0.0246	0.0240	0.0236	0.0127	0.0117	0.0110	0.0107	0.0106
Bimodal	On-target	5	0.0889	0.0822	0.0776	0.0759	0.0749	0.0279	0.0259	0.0244	0.0239	0.0235	0.0124	0.0116	0.0109	0.0107	0.0105
Bimodal	On-target	10	0.0871	0.0818	0.0771	0.0755	0.0745	0.0274	0.0258	0.0243	0.0238	0.0235	0.0122	0.0115	0.0109	0.0106	0.0105
Bimodal	On-target	15	0.0870	0.0815	0.0770	0.0754	0.0745	0.0273	0.0256	0.0243	0.0238	0.0234	0.0122	0.0114	0.0109	0.0106	0.0105
Bimodal	On-target	20	0.0867	0.0814	0.0769	0.0753	0.0744	0.0272	0.0256	0.0242	0.0237	0.0234	0.0121	0.0114	0.0108	0.0106	0.0105
Bimodal	Off by 0.5	3	0.0900	0.0838	0.0784	0.0766	0.0752	0.0284	0.0265	0.0247	0.0240	0.0237	0.0126	0.0118	0.0110	0.0108	0.0106
Bimodal	Off by 0.5	5	0.0890	0.0823	0.0778	0.0761	0.0749	0.0277	0.0260	0.0245	0.0239	0.0236	0.0125	0.0116	0.0110	0.0107	0.0105
Bimodal	Off by 0.5	10	0.0873	0.0819	0.0774	0.0757	0.0746	0.0274	0.0258	0.0243	0.0238	0.0235	0.0123	0.0115	0.0109	0.0106	0.0105
Bimodal	Off by 0.5	15	0.0866	0.0814	0.0772	0.0756	0.0746	0.0273	0.0256	0.0243	0.0238	0.0235	0.0122	0.0114	0.0109	0.0106	0.0105
Bimodal	Off by 0.5	20	0.0867	0.0813	0.0771	0.0755	0.0745	0.0272	0.0256	0.0243	0.0238	0.0235	0.0122	0.0114	0.0109	0.0106	0.0105
Bimodal	Off by 1.0	3	0.0905	0.0845	0.0788	0.0771	0.0755	0.0288	0.0265	0.0248	0.0241	0.0238	0.0129	0.0118	0.0111	0.0108	0.0106
Bimodal	Off by 1.0	5	0.0890	0.0832	0.0783	0.0764	0.0754	0.0282	0.0261	0.0247	0.0240	0.0237	0.0125	0.0117	0.0110	0.0107	0.0106
Bimodal	Off by 1.0	10	0.0879	0.0820	0.0779	0.0760	0.0750	0.0277	0.0259	0.0245	0.0239	0.0236	0.0123	0.0116	0.0110	0.0107	0.0105
Bimodal	Off by 1.0	15	0.0874	0.0821	0.0775	0.0760	0.0749	0.0275	0.0258	0.0245	0.0239	0.0236	0.0123	0.0115	0.0109	0.0107	0.0105
Bimodal	Off by 1.0	20	0.0874	0.0818	0.0776	0.0759	0.0748	0.0274	0.0257	0.0244	0.0239	0.0235	0.0122	0.0115	0.0109	0.0106	0.0105
Bimodal	Off by 2.5	3						0.0293	0.0275	0.0258	0.0252	0.0246	0.0132	0.0123	0.0116	0.0112	0.0110
Bimodal	Off by 2.5	5						0.0288	0.0272	0.0256	0.0249	0.0244	0.0129	0.0120	0.0115	0.0111	0.0109
Bimodal	Off by 2.5	10						0.0282	0.0267	0.0255	0.0247	0.0243	0.0127	0.0120	0.0114	0.0111	0.0109
Bimodal	Off by 2.5	15						0.0281	0.0267	0.0254	0.0247	0.0243	0.0125	0.0119	0.0114	0.0110	0.0109
Bimodal	Off by 2.5	20						0.0280	0.0266	0.0254	0.0247	0.0243	0.0126	0.0119	0.0113	0.0110	0.0108
Chisq(5)	On-target	3	0.0869	0.0809	0.0764	0.0741	0.0736	0.0274	0.0251	0.0240	0.0234	0.0231	0.0123	0.0113	0.0107	0.0104	0.0103
Chisq(5)	On-target	5	0.0853	0.0802	0.0757	0.0739	0.0731	0.0269	0.0249	0.0237	0.0232	0.0230	0.0120	0.0112	0.0106	0.0104	0.0102
Chisq(5)	On-target	10	0.0844	0.0792	0.0749	0.0734	0.0727	0.0265	0.0247	0.0236	0.0231	0.0229	0.0118	0.0111	0.0105	0.0103	0.0102
Chisq(5)	On-target	15	0.0838	0.0786	0.0749	0.0733	0.0727	0.0264	0.0247	0.0236	0.0230	0.0228	0.0118	0.0111	0.0105	0.0103	0.0102
Chisq(5)	On-target	20	0.0835	0.0786	0.0748	0.0733	0.0726	0.0263	0.0246	0.0235	0.0230	0.0228	0.0117	0.0110	0.0105	0.0103	0.0102
Chisq(5)	Off by 0.5	3	0.0890	0.0823	0.0769	0.0750	0.0740	0.0279	0.0257	0.0242	0.0236	0.0232	0.0126	0.0114	0.0108	0.0105	0.0104
Chisq(5)	Off by 0.5	5	0.0876	0.0808	0.0761	0.0745	0.0736	0.0272	0.0255	0.0240	0.0234	0.0231	0.0122	0.0114	0.0107	0.0105	0.0103
Chisq(5)	Off by 0.5	10	0.0859	0.0804	0.0758	0.0742	0.0732	0.0269	0.0252	0.0238	0.0233	0.0230	0.0120	0.0113	0.0107	0.0104	0.0103
Chisq(5)	Off by 0.5	15	0.0855	0.0798	0.0757	0.0741	0.0732	0.0268	0.0251	0.0238	0.0232	0.0230	0.0120	0.0112	0.0106	0.0104	0.0103
Chisq(5)	Off by 0.5	20	0.0853	0.0799	0.0757	0.0740	0.0731	0.0268	0.0251	0.0237	0.0232	0.0230	0.0120	0.0112	0.0106	0.0104	0.0103
Chisq(5)	Off by 1.0	3	0.0903	0.0838	0.0783	0.0759	0.0745	0.0284	0.0263	0.0247	0.0239	0.0234	0.0127	0.0117	0.0110	0.0106	0.0104
Chisq(5)	Off by 1.0	5	0.0888	0.0825	0.0775	0.0753	0.0739	0.0281	0.0261	0.0243	0.0237	0.0233	0.0125	0.0116	0.0109	0.0106	0.0104
Chisq(5)	Off by 1.0	10	0.0875	0.0816	0.0769	0.0751	0.0738	0.0275	0.0256	0.0242	0.0235	0.0232	0.0123	0.0115	0.0108	0.0105	0.0104
Chisq(5)	Off by 1.0	15	0.0874	0.0815	0.0767	0.0749	0.0737	0.0275	0.0256	0.0241	0.0235	0.0232	0.0123	0.0114	0.0108	0.0105	0.0103
Chisq(5)	Off by 1.0	20	0.0869	0.0812	0.0768	0.0749	0.0736	0.0273	0.0255	0.0241	0.0235	0.0232	0.0122	0.0114	0.0108	0.0105	0.0103
Chisq(5)	Off by 2.5	3						0.0320	0.0292	0.0266	0.0254	0.0247	0.0141	0.0130	0.0118	0.0113	0.0110
Chisq(5)	Off by 2.5	5						0.0309	0.0285	0.0262	0.0252	0.0246	0.0140	0.0127	0.0118	0.0113	0.0110
Chisq(5)	Off by 2.5	10						0.0307	0.0281	0.0262	0.0250	0.0244	0.0137	0.0126	0.0117	0.0112	0.0109
Chisq(5)	Off by 2.5	15						0.0306	0.0281	0.0260	0.0250	0.0244	0.0136	0.0126	0.0116	0.0112	0.0109
Chisq(5)	Off by 2.5	20						0.0303	0.0281	0.0260	0.0249	0.0244	0.0136	0.0125	0.0116	0.0112	0.0109

Table 20 Standard Error Estimated for Variance by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.1456	0.1316	0.1208	0.1156	0.1131	0.0444	0.0411	0.0374	0.0359	0.0350	0.0197	0.0183	0.0167	0.0161	0.0156
Normal	On-target	5	0.1407	0.1292	0.1192	0.1142	0.1121	0.0432	0.0399	0.0368	0.0358	0.0349	0.0190	0.0177	0.0165	0.0160	0.0155
Normal	On-target	10	0.1385	0.1270	0.1187	0.1135	0.1110	0.0422	0.0395	0.0365	0.0353	0.0346	0.0188	0.0175	0.0163	0.0157	0.0155
Normal	On-target	15	0.1368	0.1266	0.1176	0.1134	0.1108	0.0419	0.0391	0.0365	0.0352	0.0345	0.0186	0.0175	0.0162	0.0157	0.0154
Normal	On-target	20	0.1370	0.1263	0.1175	0.1130	0.1108	0.0418	0.0390	0.0364	0.0352	0.0344	0.0186	0.0174	0.0162	0.0157	0.0154
Normal	Off by 0.5	3	0.1447	0.1315	0.1221	0.1165	0.1128	0.0441	0.0406	0.0376	0.0359	0.0354	0.0197	0.0183	0.0167	0.0161	0.0157
Normal	Off by 0.5	5	0.1414	0.1297	0.1196	0.1150	0.1123	0.0430	0.0400	0.0370	0.0356	0.0350	0.0192	0.0178	0.0165	0.0159	0.0156
Normal	Off by 0.5	10	0.1383	0.1272	0.1185	0.1142	0.1113	0.0421	0.0393	0.0367	0.0352	0.0347	0.0188	0.0176	0.0164	0.0158	0.0155
Normal	Off by 0.5	15	0.1374	0.1270	0.1178	0.1134	0.1108	0.0416	0.0392	0.0365	0.0352	0.0346	0.0187	0.0175	0.0163	0.0158	0.0154
Normal	Off by 0.5	20	0.1369	0.1263	0.1177	0.1133	0.1109	0.0417	0.0391	0.0363	0.0351	0.0346	0.0186	0.0174	0.0163	0.0157	0.0154
Normal	Off by 1.0	3	0.1432	0.1347	0.1226	0.1171	0.1136	0.0444	0.0412	0.0379	0.0363	0.0354	0.0196	0.0183	0.0169	0.0162	0.0158
Normal	Off by 1.0	5	0.1423	0.1314	0.1211	0.1159	0.1125	0.0435	0.0405	0.0372	0.0359	0.0352	0.0193	0.0180	0.0167	0.0160	0.0157
Normal	Off by 1.0	10	0.1382	0.1294	0.1186	0.1144	0.1115	0.0423	0.0396	0.0369	0.0356	0.0348	0.0189	0.0177	0.0164	0.0159	0.0156
Normal	Off by 1.0	15	0.1370	0.1286	0.1185	0.1143	0.1113	0.0419	0.0394	0.0367	0.0355	0.0348	0.0188	0.0176	0.0165	0.0158	0.0155
Normal	Off by 1.0	20	0.1363	0.1282	0.1186	0.1139	0.1112	0.0419	0.0392	0.0366	0.0354	0.0347	0.0187	0.0175	0.0164	0.0158	0.0155
Normal	Off by 2.5	3						0.0456	0.0432	0.0400	0.0381	0.0369	0.0202	0.0192	0.0175	0.0170	0.0166
Normal	Off by 2.5	5						0.0441	0.0419	0.0391	0.0376	0.0366	0.0199	0.0185	0.0174	0.0169	0.0164
Normal	Off by 2.5	10						0.0433	0.0411	0.0384	0.0371	0.0361	0.0193	0.0183	0.0172	0.0167	0.0162
Normal	Off by 2.5	15						0.0431	0.0407	0.0384	0.0370	0.0360	0.0192	0.0183	0.0171	0.0166	0.0162
Normal	Off by 2.5	20						0.0429	0.0406	0.0382	0.0369	0.0359	0.0192	0.0181	0.0171	0.0165	0.0161
Uniform	On-target	3	0.1475	0.1347	0.1220	0.1169	0.1142	0.0454	0.0407	0.0380	0.0363	0.0354	0.0200	0.0185	0.0169	0.0162	0.0158
Uniform	On-target	5	0.1468	0.1325	0.1202	0.1159	0.1133	0.0437	0.0403	0.0375	0.0360	0.0351	0.0195	0.0181	0.0167	0.0160	0.0156
Uniform	On-target	10	0.1425	0.1309	0.1194	0.1148	0.1123	0.0432	0.0400	0.0369	0.0357	0.0349	0.0192	0.0179	0.0165	0.0159	0.0156
Uniform	On-target	15	0.1414	0.1295	0.1188	0.1145	0.1120	0.0428	0.0398	0.0367	0.0355	0.0348	0.0192	0.0177	0.0164	0.0158	0.0155
Uniform	On-target	20	0.1403	0.1293	0.1184	0.1144	0.1121	0.0428	0.0397	0.0367	0.0355	0.0348	0.0190	0.0177	0.0164	0.0158	0.0155
Uniform	Off by 0.5	3	0.1505	0.1369	0.1232	0.1171	0.1151	0.0461	0.0418	0.0381	0.0364	0.0356	0.0203	0.0185	0.0170	0.0163	0.0158
Uniform	Off by 0.5	5	0.1442	0.1334	0.1218	0.1160	0.1135	0.0441	0.0409	0.0375	0.0359	0.0351	0.0198	0.0182	0.0167	0.0161	0.0157
Uniform	Off by 0.5	10	0.1426	0.1308	0.1200	0.1152	0.1127	0.0437	0.0403	0.0372	0.0357	0.0350	0.0193	0.0179	0.0166	0.0159	0.0156
Uniform	Off by 0.5	15	0.1421	0.1301	0.1195	0.1151	0.1126	0.0435	0.0399	0.0370	0.0356	0.0349	0.0192	0.0179	0.0165	0.0159	0.0156
Uniform	Off by 0.5	20	0.1415	0.1300	0.1192	0.1149	0.1125	0.0432	0.0398	0.0370	0.0356	0.0348	0.0191	0.0178	0.0165	0.0158	0.0155
Uniform	Off by 1.0	3	0.1514	0.1362	0.1240	0.1189	0.1147	0.0464	0.0422	0.0387	0.0369	0.0360	0.0204	0.0189	0.0171	0.0165	0.0160
Uniform	Off by 1.0	5	0.1486	0.1350	0.1219	0.1175	0.1143	0.0450	0.0414	0.0379	0.0365	0.0356	0.0200	0.0185	0.0169	0.0162	0.0158
Uniform	Off by 1.0	10	0.1452	0.1324	0.1212	0.1166	0.1134	0.0441	0.0408	0.0376	0.0362	0.0353	0.0195	0.0182	0.0168	0.0161	0.0157
Uniform	Off by 1.0	15	0.1430	0.1320	0.1208	0.1163	0.1133	0.0437	0.0405	0.0374	0.0361	0.0352	0.0194	0.0180	0.0167	0.0161	0.0157
Uniform	Off by 1.0	20	0.1435	0.1316	0.1203	0.1159	0.1130	0.0434	0.0404	0.0374	0.0359	0.0352	0.0193	0.0180	0.0166	0.0160	0.0157
Uniform	Off by 2.5	3						0.0470	0.0444	0.0414	0.0396	0.0385	0.0208	0.0197	0.0186	0.0177	0.0173
Uniform	Off by 2.5	5						0.0453	0.0436	0.0407	0.0391	0.0382	0.0204	0.0194	0.0183	0.0175	0.0170
Uniform	Off by 2.5	10						0.0441	0.0427	0.0403	0.0388	0.0377	0.0197	0.0190	0.0180	0.0173	0.0168
Uniform	Off by 2.5	15						0.0442	0.0424	0.0402	0.0387	0.0376	0.0197	0.0190	0.0179	0.0173	0.0168
Uniform	Off by 2.5	20						0.0438	0.0422	0.0400	0.0386	0.0376	0.0196	0.0190	0.0179	0.0172	0.0167
Bimodal	On-target	3	0.1460	0.1328	0.1216	0.1156	0.1124	0.0453	0.0412	0.0375	0.0360	0.0354	0.0200	0.0186	0.0170	0.0163	0.0158

Appendix 4

Standard Error Estimates of Simulation

Bimodal	On-target	5	0.1425	0.1319	0.1194	0.1149	0.1123	0.0440	0.0408	0.0371	0.0359	0.0349	0.0197	0.0182	0.0167	0.0160	0.0156
Bimodal	On-target	10	0.1400	0.1292	0.1181	0.1139	0.1114	0.0431	0.0399	0.0368	0.0356	0.0347	0.0192	0.0178	0.0165	0.0159	0.0155
Bimodal	On-target	15	0.1394	0.1288	0.1179	0.1136	0.1109	0.0427	0.0399	0.0367	0.0354	0.0346	0.0192	0.0178	0.0164	0.0158	0.0155
Bimodal	On-target	20	0.1383	0.1286	0.1177	0.1134	0.1108	0.0427	0.0398	0.0368	0.0354	0.0346	0.0191	0.0178	0.0164	0.0158	0.0155
Bimodal	Off by 0.5	3	0.1450	0.1333	0.1227	0.1172	0.1125	0.0444	0.0412	0.0380	0.0364	0.0353	0.0202	0.0187	0.0171	0.0162	0.0158
Bimodal	Off by 0.5	5	0.1414	0.1310	0.1203	0.1155	0.1120	0.0440	0.0407	0.0374	0.0359	0.0352	0.0197	0.0182	0.0168	0.0161	0.0157
Bimodal	Off by 0.5	10	0.1405	0.1298	0.1193	0.1144	0.1115	0.0430	0.0404	0.0371	0.0356	0.0349	0.0193	0.0180	0.0166	0.0159	0.0156
Bimodal	Off by 0.5	15	0.1393	0.1285	0.1181	0.1141	0.1113	0.0429	0.0402	0.0370	0.0355	0.0348	0.0192	0.0179	0.0165	0.0159	0.0155
Bimodal	Off by 0.5	20	0.1390	0.1285	0.1182	0.1138	0.1110	0.0428	0.0399	0.0370	0.0356	0.0347	0.0191	0.0178	0.0165	0.0158	0.0155
Bimodal	Off by 1.0	3	0.1504	0.1355	0.1233	0.1176	0.1148	0.0458	0.0423	0.0385	0.0367	0.0357	0.0203	0.0190	0.0171	0.0164	0.0159
Bimodal	Off by 1.0	5	0.1458	0.1336	0.1217	0.1168	0.1134	0.0442	0.0414	0.0381	0.0363	0.0354	0.0198	0.0184	0.0170	0.0162	0.0158
Bimodal	Off by 1.0	10	0.1427	0.1309	0.1205	0.1155	0.1127	0.0436	0.0406	0.0376	0.0359	0.0351	0.0194	0.0182	0.0167	0.0161	0.0157
Bimodal	Off by 1.0	15	0.1420	0.1305	0.1195	0.1153	0.1123	0.0435	0.0406	0.0375	0.0359	0.0351	0.0194	0.0181	0.0167	0.0161	0.0156
Bimodal	Off by 1.0	20	0.1412	0.1296	0.1194	0.1149	0.1123	0.0434	0.0405	0.0373	0.0359	0.0351	0.0194	0.0180	0.0166	0.0160	0.0156
Bimodal	Off by 2.5	3						0.0468	0.0442	0.0415	0.0402	0.0385	0.0205	0.0200	0.0187	0.0179	0.0173
Bimodal	Off by 2.5	5						0.0457	0.0438	0.0411	0.0395	0.0380	0.0202	0.0195	0.0185	0.0176	0.0170
Bimodal	Off by 2.5	10						0.0446	0.0429	0.0408	0.0390	0.0377	0.0199	0.0192	0.0180	0.0173	0.0168
Bimodal	Off by 2.5	15						0.0442	0.0424	0.0403	0.0388	0.0375	0.0197	0.0191	0.0181	0.0173	0.0168
Bimodal	Off by 2.5	20						0.0441	0.0425	0.0402	0.0386	0.0374	0.0196	0.0189	0.0179	0.0173	0.0167
Chisq(5)	On-target	3	0.1384	0.1265	0.1162	0.1109	0.1080	0.0417	0.0389	0.0357	0.0342	0.0335	0.0184	0.0174	0.0160	0.0153	0.0150
Chisq(5)	On-target	5	0.1348	0.1238	0.1142	0.1094	0.1077	0.0409	0.0378	0.0350	0.0341	0.0332	0.0182	0.0168	0.0157	0.0152	0.0148
Chisq(5)	On-target	10	0.1299	0.1213	0.1126	0.1087	0.1068	0.0399	0.0371	0.0348	0.0337	0.0330	0.0177	0.0167	0.0156	0.0150	0.0147
Chisq(5)	On-target	15	0.1293	0.1204	0.1121	0.1083	0.1066	0.0396	0.0369	0.0346	0.0334	0.0330	0.0176	0.0165	0.0155	0.0150	0.0147
Chisq(5)	On-target	20	0.1293	0.1206	0.1119	0.1081	0.1065	0.0396	0.0366	0.0345	0.0334	0.0329	0.0176	0.0165	0.0155	0.0149	0.0147
Chisq(5)	Off by 0.5	3	0.1416	0.1287	0.1165	0.1119	0.1097	0.0435	0.0398	0.0358	0.0348	0.0337	0.0193	0.0175	0.0162	0.0155	0.0152
Chisq(5)	Off by 0.5	5	0.1383	0.1261	0.1160	0.1112	0.1087	0.0421	0.0391	0.0357	0.0343	0.0336	0.0188	0.0173	0.0160	0.0154	0.0150
Chisq(5)	Off by 0.5	10	0.1370	0.1250	0.1141	0.1104	0.1077	0.0414	0.0383	0.0354	0.0340	0.0333	0.0185	0.0171	0.0158	0.0152	0.0148
Chisq(5)	Off by 0.5	15	0.1357	0.1240	0.1139	0.1100	0.1076	0.0412	0.0381	0.0352	0.0339	0.0333	0.0184	0.0170	0.0157	0.0152	0.0148
Chisq(5)	Off by 0.5	20	0.1353	0.1235	0.1140	0.1098	0.1075	0.0409	0.0380	0.0351	0.0338	0.0332	0.0183	0.0169	0.0157	0.0151	0.0148
Chisq(5)	Off by 1.0	3	0.1484	0.1343	0.1200	0.1155	0.1109	0.0449	0.0409	0.0369	0.0354	0.0343	0.0200	0.0185	0.0167	0.0157	0.0153
Chisq(5)	Off by 1.0	5	0.1443	0.1312	0.1186	0.1134	0.1098	0.0446	0.0405	0.0367	0.0349	0.0341	0.0197	0.0181	0.0164	0.0156	0.0152
Chisq(5)	Off by 1.0	10	0.1424	0.1291	0.1176	0.1125	0.1091	0.0430	0.0399	0.0364	0.0348	0.0339	0.0195	0.0178	0.0162	0.0155	0.0151
Chisq(5)	Off by 1.0	15	0.1419	0.1283	0.1172	0.1124	0.1086	0.0428	0.0397	0.0360	0.0347	0.0337	0.0193	0.0176	0.0161	0.0154	0.0150
Chisq(5)	Off by 1.0	20	0.1407	0.1281	0.1168	0.1120	0.1088	0.0429	0.0395	0.0361	0.0346	0.0337	0.0192	0.0176	0.0161	0.0154	0.0150
Chisq(5)	Off by 2.5	3						0.0556	0.0500	0.0430	0.0398	0.0380	0.0248	0.0219	0.0193	0.0178	0.0170
Chisq(5)	Off by 2.5	5						0.0536	0.0482	0.0423	0.0392	0.0377	0.0243	0.0217	0.0190	0.0176	0.0168
Chisq(5)	Off by 2.5	10						0.0530	0.0477	0.0420	0.0388	0.0374	0.0236	0.0212	0.0187	0.0174	0.0166
Chisq(5)	Off by 2.5	15						0.0524	0.0470	0.0417	0.0388	0.0372	0.0235	0.0210	0.0186	0.0173	0.0166
Chisq(5)	Off by 2.5	20						0.0523	0.0470	0.0416	0.0387	0.0371	0.0234	0.0210	0.0186	0.0173	0.0166

Table 21 Standard Error Estimated for Percentage in Band by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.0390	0.0371	0.0346	0.0336	0.0331	0.0122	0.0117	0.0109	0.0107	0.0105	0.0055	0.0052	0.0049	0.0048	0.0047
Normal	On-target	5	0.0372	0.0360	0.0341	0.0334	0.0328	0.0119	0.0113	0.0108	0.0106	0.0103	0.0053	0.0051	0.0048	0.0047	0.0046
Normal	On-target	10	0.0371	0.0353	0.0337	0.0329	0.0324	0.0117	0.0112	0.0106	0.0104	0.0102	0.0052	0.0050	0.0048	0.0047	0.0046
Normal	On-target	15	0.0368	0.0351	0.0335	0.0328	0.0322	0.0117	0.0111	0.0106	0.0104	0.0102	0.0052	0.0050	0.0047	0.0046	0.0046
Normal	On-target	20	0.0365	0.0351	0.0336	0.0327	0.0322	0.0116	0.0111	0.0106	0.0103	0.0102	0.0052	0.0049	0.0047	0.0046	0.0046
Normal	Off by 0.5	3	0.0384	0.0361	0.0345	0.0336	0.0328	0.0121	0.0115	0.0109	0.0106	0.0105	0.0054	0.0052	0.0049	0.0048	0.0047
Normal	Off by 0.5	5	0.0375	0.0357	0.0340	0.0333	0.0325	0.0118	0.0113	0.0108	0.0106	0.0103	0.0053	0.0050	0.0048	0.0047	0.0046
Normal	Off by 0.5	10	0.0367	0.0352	0.0336	0.0326	0.0323	0.0118	0.0112	0.0106	0.0104	0.0102	0.0052	0.0050	0.0047	0.0047	0.0046
Normal	Off by 0.5	15	0.0366	0.0349	0.0334	0.0327	0.0321	0.0117	0.0111	0.0106	0.0103	0.0102	0.0052	0.0050	0.0047	0.0046	0.0045
Normal	Off by 0.5	20	0.0365	0.0348	0.0334	0.0326	0.0321	0.0116	0.0111	0.0105	0.0103	0.0102	0.0052	0.0049	0.0047	0.0046	0.0045
Normal	Off by 1.0	3	0.0383	0.0365	0.0343	0.0335	0.0328	0.0122	0.0115	0.0109	0.0107	0.0105	0.0055	0.0052	0.0049	0.0048	0.0047
Normal	Off by 1.0	5	0.0373	0.0358	0.0337	0.0330	0.0324	0.0119	0.0114	0.0107	0.0104	0.0102	0.0054	0.0051	0.0048	0.0047	0.0046
Normal	Off by 1.0	10	0.0367	0.0350	0.0333	0.0325	0.0321	0.0117	0.0111	0.0106	0.0104	0.0102	0.0052	0.0050	0.0048	0.0046	0.0046
Normal	Off by 1.0	15	0.0363	0.0348	0.0332	0.0325	0.0320	0.0116	0.0111	0.0106	0.0103	0.0101	0.0052	0.0050	0.0047	0.0046	0.0045
Normal	Off by 1.0	20	0.0361	0.0348	0.0332	0.0323	0.0319	0.0116	0.0111	0.0105	0.0103	0.0101	0.0052	0.0049	0.0047	0.0046	0.0045
Normal	Off by 2.5	3						0.0123	0.0116	0.0109	0.0107	0.0106	0.0054	0.0051	0.0049	0.0048	0.0047
Normal	Off by 2.5	5						0.0119	0.0114	0.0109	0.0106	0.0103	0.0053	0.0051	0.0048	0.0047	0.0046
Normal	Off by 2.5	10						0.0117	0.0112	0.0107	0.0104	0.0102	0.0052	0.0050	0.0048	0.0047	0.0046
Normal	Off by 2.5	15						0.0116	0.0112	0.0106	0.0104	0.0102	0.0052	0.0050	0.0048	0.0046	0.0046
Normal	Off by 2.5	20						0.0116	0.0111	0.0106	0.0104	0.0102	0.0052	0.0050	0.0048	0.0046	0.0046
Uniform	On-target	3	0.0355	0.0342	0.0336	0.0329	0.0325	0.0111	0.0108	0.0104	0.0103	0.0102	0.0050	0.0048	0.0047	0.0046	0.0045
Uniform	On-target	5	0.0346	0.0337	0.0328	0.0324	0.0319	0.0109	0.0105	0.0103	0.0102	0.0101	0.0049	0.0047	0.0046	0.0045	0.0045
Uniform	On-target	10	0.0336	0.0329	0.0323	0.0319	0.0317	0.0107	0.0103	0.0102	0.0100	0.0100	0.0048	0.0046	0.0045	0.0045	0.0045
Uniform	On-target	15	0.0336	0.0328	0.0320	0.0318	0.0315	0.0106	0.0103	0.0101	0.0100	0.0099	0.0047	0.0046	0.0045	0.0045	0.0044
Uniform	On-target	20	0.0333	0.0326	0.0320	0.0317	0.0315	0.0106	0.0103	0.0100	0.0100	0.0099	0.0047	0.0046	0.0045	0.0045	0.0044
Uniform	Off by 0.5	3	0.0351	0.0341	0.0334	0.0323	0.0325	0.0111	0.0107	0.0104	0.0103	0.0102	0.0049	0.0048	0.0047	0.0046	0.0046
Uniform	Off by 0.5	5	0.0340	0.0337	0.0324	0.0323	0.0319	0.0108	0.0105	0.0103	0.0101	0.0101	0.0048	0.0047	0.0046	0.0045	0.0045
Uniform	Off by 0.5	10	0.0337	0.0329	0.0320	0.0317	0.0314	0.0106	0.0104	0.0101	0.0100	0.0099	0.0048	0.0046	0.0046	0.0045	0.0045
Uniform	Off by 0.5	15	0.0333	0.0325	0.0320	0.0315	0.0313	0.0106	0.0103	0.0101	0.0100	0.0099	0.0047	0.0046	0.0045	0.0044	0.0044
Uniform	Off by 0.5	20	0.0331	0.0325	0.0319	0.0315	0.0313	0.0105	0.0102	0.0101	0.0100	0.0099	0.0047	0.0046	0.0045	0.0044	0.0044
Uniform	Off by 1.0	3	0.0352	0.0343	0.0329	0.0326	0.0322	0.0111	0.0108	0.0105	0.0103	0.0102	0.0050	0.0048	0.0047	0.0046	0.0045
Uniform	Off by 1.0	5	0.0342	0.0333	0.0324	0.0321	0.0317	0.0108	0.0105	0.0102	0.0101	0.0101	0.0048	0.0047	0.0046	0.0045	0.0045
Uniform	Off by 1.0	10	0.0335	0.0326	0.0318	0.0315	0.0313	0.0106	0.0103	0.0101	0.0100	0.0099	0.0048	0.0046	0.0045	0.0045	0.0044
Uniform	Off by 1.0	15	0.0332	0.0324	0.0317	0.0314	0.0312	0.0106	0.0103	0.0101	0.0100	0.0099	0.0048	0.0046	0.0045	0.0045	0.0044
Uniform	Off by 1.0	20	0.0331	0.0322	0.0317	0.0314	0.0312	0.0105	0.0103	0.0100	0.0099	0.0099	0.0047	0.0046	0.0045	0.0044	0.0044
Uniform	Off by 2.5	3						0.0110	0.0108	0.0104	0.0103	0.0102	0.0049	0.0047	0.0046	0.0046	0.0045
Uniform	Off by 2.5	5						0.0108	0.0105	0.0102	0.0101	0.0100	0.0048	0.0047	0.0046	0.0046	0.0045
Uniform	Off by 2.5	10						0.0106	0.0103	0.0101	0.0100	0.0099	0.0047	0.0046	0.0045	0.0045	0.0044
Uniform	Off by 2.5	15						0.0105	0.0103	0.0100	0.0100	0.0099	0.0047	0.0046	0.0045	0.0045	0.0044
Uniform	Off by 2.5	20						0.0104	0.0102	0.0100	0.0099	0.0099	0.0047	0.0046	0.0045	0.0045	0.0044
Bimodal	On-target	3	0.0368	0.0363	0.0353	0.0352	0.0347	0.0118	0.0115	0.0111	0.0110	0.0109	0.0052	0.0051	0.0050	0.0050	0.0049

Appendix 4

Standard Error Estimates of Simulation

Bimodal	On-target	5	0.0366	0.0353	0.0345	0.0344	0.0342	0.0115	0.0112	0.0110	0.0109	0.0109	0.0052	0.0050	0.0049	0.0049	0.0048
Bimodal	On-target	10	0.0355	0.0347	0.0343	0.0339	0.0338	0.0113	0.0110	0.0108	0.0107	0.0107	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	On-target	15	0.0352	0.0347	0.0341	0.0339	0.0337	0.0112	0.0109	0.0108	0.0107	0.0107	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	On-target	20	0.0352	0.0345	0.0340	0.0339	0.0336	0.0112	0.0109	0.0107	0.0106	0.0106	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	Off by 0.5	3	0.0367	0.0357	0.0356	0.0349	0.0346	0.0117	0.0115	0.0111	0.0111	0.0109	0.0052	0.0051	0.0050	0.0050	0.0048
Bimodal	Off by 0.5	5	0.0358	0.0351	0.0346	0.0344	0.0341	0.0115	0.0111	0.0109	0.0109	0.0108	0.0051	0.0050	0.0049	0.0049	0.0048
Bimodal	Off by 0.5	10	0.0354	0.0347	0.0341	0.0338	0.0337	0.0112	0.0110	0.0108	0.0107	0.0106	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	Off by 0.5	15	0.0349	0.0343	0.0338	0.0336	0.0335	0.0112	0.0109	0.0108	0.0107	0.0106	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	Off by 0.5	20	0.0348	0.0342	0.0338	0.0336	0.0335	0.0112	0.0109	0.0107	0.0107	0.0106	0.0050	0.0049	0.0048	0.0048	0.0047
Bimodal	Off by 1.0	3	0.0364	0.0354	0.0349	0.0346	0.0345	0.0118	0.0114	0.0111	0.0110	0.0110	0.0053	0.0051	0.0050	0.0050	0.0049
Bimodal	Off by 1.0	5	0.0357	0.0350	0.0344	0.0343	0.0340	0.0115	0.0112	0.0111	0.0108	0.0107	0.0051	0.0050	0.0049	0.0049	0.0049
Bimodal	Off by 1.0	10	0.0352	0.0343	0.0339	0.0338	0.0334	0.0113	0.0109	0.0107	0.0107	0.0106	0.0050	0.0049	0.0048	0.0048	0.0048
Bimodal	Off by 1.0	15	0.0349	0.0342	0.0336	0.0335	0.0334	0.0111	0.0109	0.0107	0.0107	0.0106	0.0050	0.0049	0.0048	0.0048	0.0047
Bimodal	Off by 1.0	20	0.0348	0.0341	0.0336	0.0335	0.0333	0.0111	0.0109	0.0107	0.0106	0.0106	0.0050	0.0049	0.0048	0.0048	0.0047
Bimodal	Off by 2.5	3						0.0117	0.0113	0.0111	0.0109	0.0109	0.0053	0.0050	0.0050	0.0049	0.0049
Bimodal	Off by 2.5	5						0.0114	0.0112	0.0109	0.0108	0.0107	0.0051	0.0050	0.0049	0.0048	0.0048
Bimodal	Off by 2.5	10						0.0112	0.0109	0.0108	0.0106	0.0106	0.0051	0.0049	0.0048	0.0048	0.0048
Bimodal	Off by 2.5	15						0.0111	0.0109	0.0107	0.0106	0.0106	0.0050	0.0049	0.0048	0.0048	0.0047
Bimodal	Off by 2.5	20						0.0111	0.0109	0.0107	0.0106	0.0105	0.0050	0.0049	0.0048	0.0048	0.0047
Chisq(5)	On-target	3	0.0393	0.0370	0.0344	0.0330	0.0325	0.0126	0.0116	0.0109	0.0105	0.0103	0.0057	0.0052	0.0049	0.0047	0.0046
Chisq(5)	On-target	5	0.0383	0.0364	0.0338	0.0330	0.0321	0.0123	0.0114	0.0106	0.0103	0.0102	0.0055	0.0051	0.0048	0.0046	0.0046
Chisq(5)	On-target	10	0.0380	0.0357	0.0335	0.0324	0.0320	0.0121	0.0112	0.0106	0.0103	0.0101	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	On-target	15	0.0376	0.0353	0.0334	0.0323	0.0318	0.0120	0.0112	0.0105	0.0102	0.0100	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	On-target	20	0.0376	0.0353	0.0332	0.0323	0.0317	0.0119	0.0112	0.0105	0.0102	0.0100	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 0.5	3	0.0395	0.0371	0.0344	0.0332	0.0326	0.0125	0.0117	0.0108	0.0105	0.0102	0.0057	0.0051	0.0049	0.0047	0.0046
Chisq(5)	Off by 0.5	5	0.0392	0.0360	0.0338	0.0327	0.0320	0.0123	0.0114	0.0107	0.0104	0.0102	0.0055	0.0051	0.0048	0.0046	0.0046
Chisq(5)	Off by 0.5	10	0.0380	0.0356	0.0334	0.0323	0.0318	0.0121	0.0113	0.0106	0.0102	0.0101	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 0.5	15	0.0378	0.0353	0.0332	0.0322	0.0316	0.0120	0.0112	0.0105	0.0102	0.0100	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 0.5	20	0.0376	0.0353	0.0332	0.0322	0.0317	0.0120	0.0112	0.0105	0.0102	0.0100	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 1.0	3	0.0396	0.0366	0.0338	0.0332	0.0324	0.0126	0.0117	0.0109	0.0105	0.0103	0.0057	0.0052	0.0049	0.0047	0.0046
Chisq(5)	Off by 1.0	5	0.0386	0.0358	0.0336	0.0327	0.0320	0.0124	0.0114	0.0107	0.0104	0.0102	0.0055	0.0051	0.0048	0.0047	0.0046
Chisq(5)	Off by 1.0	10	0.0378	0.0354	0.0333	0.0323	0.0317	0.0121	0.0113	0.0106	0.0103	0.0101	0.0055	0.0051	0.0047	0.0046	0.0045
Chisq(5)	Off by 1.0	15	0.0377	0.0352	0.0332	0.0322	0.0316	0.0121	0.0112	0.0106	0.0102	0.0101	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 1.0	20	0.0377	0.0353	0.0331	0.0322	0.0316	0.0120	0.0112	0.0106	0.0102	0.0100	0.0054	0.0050	0.0047	0.0046	0.0045
Chisq(5)	Off by 2.5	3						0.0130	0.0121	0.0111	0.0106	0.0104	0.0058	0.0055	0.0050	0.0048	0.0047
Chisq(5)	Off by 2.5	5						0.0126	0.0117	0.0109	0.0105	0.0103	0.0057	0.0053	0.0049	0.0047	0.0046
Chisq(5)	Off by 2.5	10						0.0123	0.0116	0.0108	0.0105	0.0102	0.0056	0.0052	0.0048	0.0047	0.0046
Chisq(5)	Off by 2.5	15						0.0123	0.0115	0.0107	0.0104	0.0102	0.0055	0.0052	0.0048	0.0047	0.0046
Chisq(5)	Off by 2.5	20						0.0123	0.0115	0.0107	0.0104	0.0102	0.0055	0.0052	0.0048	0.0047	0.0046

Table 22 Standard Error Estimated for 90th Percentile by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.1715	0.1625	0.1523	0.1481	0.1455	0.0537	0.0512	0.0487	0.0470	0.0461	0.0245	0.0233	0.0218	0.0213	0.0210
Normal	On-target	5	0.1658	0.1597	0.1494	0.1449	0.1429	0.0532	0.0502	0.0473	0.0463	0.0454	0.0238	0.0228	0.0215	0.0208	0.0205
Normal	On-target	10	0.1625	0.1546	0.1478	0.1435	0.1419	0.0517	0.0492	0.0469	0.0456	0.0448	0.0233	0.0222	0.0210	0.0207	0.0201
Normal	On-target	15	0.1609	0.1543	0.1475	0.1434	0.1414	0.0512	0.0489	0.0466	0.0454	0.0447	0.0232	0.0222	0.0210	0.0205	0.0202
Normal	On-target	20	0.1604	0.1538	0.1473	0.1431	0.1407	0.0509	0.0488	0.0464	0.0453	0.0445	0.0230	0.0220	0.0209	0.0205	0.0201
Normal	Off by 0.5	3	0.1687	0.1601	0.1532	0.1487	0.1443	0.0540	0.0519	0.0484	0.0470	0.0460	0.0245	0.0234	0.0221	0.0213	0.0207
Normal	Off by 0.5	5	0.1662	0.1577	0.1489	0.1450	0.1429	0.0520	0.0500	0.0471	0.0459	0.0453	0.0237	0.0224	0.0214	0.0209	0.0205
Normal	Off by 0.5	10	0.1618	0.1550	0.1481	0.1439	0.1406	0.0519	0.0491	0.0469	0.0455	0.0448	0.0233	0.0222	0.0210	0.0207	0.0203
Normal	Off by 0.5	15	0.1609	0.1543	0.1472	0.1425	0.1406	0.0511	0.0489	0.0465	0.0454	0.0446	0.0232	0.0220	0.0210	0.0206	0.0201
Normal	Off by 0.5	20	0.1600	0.1529	0.1467	0.1423	0.1400	0.0509	0.0486	0.0464	0.0452	0.0444	0.0230	0.0219	0.0209	0.0205	0.0201
Normal	Off by 1.0	3	0.1679	0.1595	0.1519	0.1470	0.1442	0.0534	0.0512	0.0483	0.0466	0.0457	0.0243	0.0231	0.0217	0.0211	0.0208
Normal	Off by 1.0	5	0.1635	0.1571	0.1485	0.1454	0.1422	0.0523	0.0501	0.0473	0.0459	0.0454	0.0238	0.0225	0.0214	0.0207	0.0204
Normal	Off by 1.0	10	0.1619	0.1545	0.1474	0.1430	0.1404	0.0513	0.0491	0.0468	0.0455	0.0448	0.0230	0.0222	0.0210	0.0206	0.0202
Normal	Off by 1.0	15	0.1604	0.1541	0.1465	0.1428	0.1407	0.0510	0.0486	0.0465	0.0452	0.0444	0.0231	0.0220	0.0210	0.0205	0.0201
Normal	Off by 1.0	20	0.1597	0.1541	0.1465	0.1423	0.1404	0.0510	0.0485	0.0463	0.0451	0.0443	0.0229	0.0220	0.0209	0.0205	0.0200
Normal	Off by 2.5	3						0.0542	0.0512	0.0483	0.0470	0.0456	0.0240	0.0230	0.0222	0.0210	0.0209
Normal	Off by 2.5	5						0.0527	0.0498	0.0474	0.0463	0.0453	0.0237	0.0226	0.0216	0.0209	0.0205
Normal	Off by 2.5	10						0.0513	0.0494	0.0468	0.0459	0.0449	0.0232	0.0223	0.0211	0.0206	0.0202
Normal	Off by 2.5	15						0.0513	0.0488	0.0468	0.0456	0.0445	0.0231	0.0221	0.0211	0.0204	0.0202
Normal	Off by 2.5	20						0.0510	0.0488	0.0464	0.0453	0.0444	0.0231	0.0221	0.0210	0.0204	0.0201
Uniform	On-target	3	0.1713	0.1651	0.1541	0.1498	0.1479	0.0542	0.0520	0.0487	0.0476	0.0468	0.0246	0.0236	0.0222	0.0216	0.0212
Uniform	On-target	5	0.1702	0.1604	0.1520	0.1493	0.1457	0.0531	0.0506	0.0479	0.0471	0.0465	0.0241	0.0230	0.0218	0.0214	0.0210
Uniform	On-target	10	0.1646	0.1581	0.1501	0.1464	0.1447	0.0522	0.0497	0.0474	0.0465	0.0460	0.0235	0.0225	0.0216	0.0210	0.0207
Uniform	On-target	15	0.1648	0.1572	0.1493	0.1465	0.1441	0.0516	0.0498	0.0470	0.0462	0.0457	0.0235	0.0225	0.0214	0.0210	0.0207
Uniform	On-target	20	0.1636	0.1564	0.1491	0.1461	0.1440	0.0517	0.0494	0.0470	0.0463	0.0458	0.0234	0.0224	0.0214	0.0209	0.0206
Uniform	Off by 0.5	3	0.1702	0.1659	0.1542	0.1484	0.1479	0.0545	0.0513	0.0481	0.0472	0.0467	0.0247	0.0235	0.0219	0.0217	0.0212
Uniform	Off by 0.5	5	0.1664	0.1606	0.1501	0.1474	0.1462	0.0531	0.0507	0.0480	0.0468	0.0462	0.0241	0.0231	0.0217	0.0212	0.0209
Uniform	Off by 0.5	10	0.1652	0.1576	0.1497	0.1461	0.1447	0.0523	0.0498	0.0472	0.0463	0.0459	0.0236	0.0225	0.0214	0.0211	0.0207
Uniform	Off by 0.5	15	0.1645	0.1568	0.1487	0.1457	0.1444	0.0520	0.0493	0.0471	0.0461	0.0456	0.0236	0.0224	0.0213	0.0209	0.0207
Uniform	Off by 0.5	20	0.1635	0.1567	0.1484	0.1455	0.1441	0.0515	0.0492	0.0470	0.0459	0.0455	0.0235	0.0224	0.0213	0.0208	0.0206
Uniform	Off by 1.0	3	0.1735	0.1626	0.1538	0.1493	0.1470	0.0548	0.0514	0.0487	0.0473	0.0468	0.0247	0.0230	0.0222	0.0215	0.0211
Uniform	Off by 1.0	5	0.1688	0.1594	0.1514	0.1473	0.1453	0.0530	0.0505	0.0475	0.0466	0.0463	0.0241	0.0229	0.0217	0.0212	0.0207
Uniform	Off by 1.0	10	0.1654	0.1578	0.1490	0.1460	0.1448	0.0520	0.0495	0.0473	0.0463	0.0458	0.0236	0.0224	0.0214	0.0209	0.0207
Uniform	Off by 1.0	15	0.1639	0.1563	0.1488	0.1451	0.1437	0.0517	0.0492	0.0470	0.0461	0.0456	0.0235	0.0223	0.0213	0.0208	0.0206
Uniform	Off by 1.0	20	0.1638	0.1559	0.1483	0.1450	0.1436	0.0516	0.0493	0.0469	0.0460	0.0455	0.0233	0.0223	0.0212	0.0208	0.0206
Uniform	Off by 2.5	3						0.0539	0.0520	0.0484	0.0471	0.0470	0.0248	0.0231	0.0219	0.0214	0.0214
Uniform	Off by 2.5	5						0.0527	0.0504	0.0481	0.0466	0.0461	0.0239	0.0228	0.0217	0.0212	0.0209
Uniform	Off by 2.5	10						0.0520	0.0496	0.0474	0.0462	0.0458	0.0232	0.0223	0.0213	0.0210	0.0207
Uniform	Off by 2.5	15						0.0517	0.0493	0.0470	0.0461	0.0455	0.0234	0.0223	0.0213	0.0209	0.0206
Uniform	Off by 2.5	20						0.0512	0.0493	0.0470	0.0460	0.0455	0.0233	0.0222	0.0213	0.0208	0.0205
Bimodal	On-target	3	0.1738	0.1608	0.1522	0.1468	0.1422	0.0545	0.0514	0.0477	0.0457	0.0448	0.0248	0.0233	0.0219	0.0210	0.0204

Appendix 4

Standard Error Estimates of Simulation

Bimodal	On-target	5	0.1685	0.1598	0.1487	0.1449	0.1407	0.0532	0.0505	0.0469	0.0455	0.0443	0.0240	0.0230	0.0215	0.0207	0.0200
Bimodal	On-target	10	0.1650	0.1569	0.1472	0.1432	0.1386	0.0520	0.0496	0.0465	0.0451	0.0440	0.0236	0.0226	0.0211	0.0205	0.0200
Bimodal	On-target	15	0.1637	0.1570	0.1471	0.1424	0.1386	0.0516	0.0494	0.0464	0.0449	0.0438	0.0234	0.0224	0.0211	0.0204	0.0198
Bimodal	On-target	20	0.1632	0.1564	0.1466	0.1423	0.1385	0.0518	0.0495	0.0463	0.0448	0.0437	0.0234	0.0223	0.0210	0.0204	0.0198
Bimodal	Off by 0.5	3	0.1715	0.1617	0.1506	0.1451	0.1414	0.0535	0.0521	0.0474	0.0459	0.0448	0.0248	0.0234	0.0217	0.0206	0.0203
Bimodal	Off by 0.5	5	0.1670	0.1589	0.1484	0.1438	0.1394	0.0526	0.0499	0.0470	0.0452	0.0444	0.0242	0.0227	0.0213	0.0206	0.0200
Bimodal	Off by 0.5	10	0.1648	0.1564	0.1467	0.1423	0.1383	0.0519	0.0495	0.0462	0.0448	0.0438	0.0234	0.0224	0.0211	0.0203	0.0198
Bimodal	Off by 0.5	15	0.1632	0.1548	0.1465	0.1418	0.1381	0.0516	0.0491	0.0462	0.0448	0.0438	0.0234	0.0222	0.0209	0.0203	0.0198
Bimodal	Off by 0.5	20	0.1628	0.1549	0.1458	0.1414	0.1376	0.0515	0.0490	0.0461	0.0447	0.0437	0.0233	0.0222	0.0209	0.0202	0.0197
Bimodal	Off by 1.0	3	0.1711	0.1610	0.1500	0.1448	0.1411	0.0540	0.0512	0.0475	0.0458	0.0448	0.0247	0.0236	0.0217	0.0206	0.0204
Bimodal	Off by 1.0	5	0.1676	0.1574	0.1473	0.1427	0.1393	0.0533	0.0507	0.0471	0.0453	0.0441	0.0240	0.0226	0.0213	0.0205	0.0201
Bimodal	Off by 1.0	10	0.1643	0.1554	0.1466	0.1419	0.1383	0.0520	0.0493	0.0464	0.0447	0.0437	0.0235	0.0224	0.0210	0.0202	0.0198
Bimodal	Off by 1.0	15	0.1635	0.1549	0.1457	0.1419	0.1379	0.0517	0.0491	0.0461	0.0446	0.0437	0.0234	0.0222	0.0208	0.0201	0.0197
Bimodal	Off by 1.0	20	0.1629	0.1550	0.1456	0.1414	0.1375	0.0517	0.0490	0.0459	0.0445	0.0435	0.0234	0.0221	0.0209	0.0201	0.0197
Bimodal	Off by 2.5	3						0.0544	0.0513	0.0480	0.0459	0.0448	0.0243	0.0230	0.0215	0.0208	0.0203
Bimodal	Off by 2.5	5						0.0531	0.0502	0.0471	0.0452	0.0442	0.0240	0.0229	0.0215	0.0205	0.0201
Bimodal	Off by 2.5	10						0.0520	0.0497	0.0468	0.0451	0.0440	0.0234	0.0224	0.0210	0.0203	0.0198
Bimodal	Off by 2.5	15						0.0513	0.0492	0.0465	0.0449	0.0437	0.0232	0.0222	0.0210	0.0202	0.0198
Bimodal	Off by 2.5	20						0.0516	0.0489	0.0463	0.0448	0.0437	0.0231	0.0222	0.0209	0.0202	0.0198
Chisq(5)	On-target	3	0.1717	0.1694	0.1663	0.1645	0.1612	0.0540	0.0531	0.0523	0.0515	0.0512	0.0247	0.0241	0.0238	0.0235	0.0231
Chisq(5)	On-target	5	0.1658	0.1654	0.1616	0.1613	0.1594	0.0524	0.0513	0.0513	0.0506	0.0503	0.0239	0.0235	0.0233	0.0229	0.0229
Chisq(5)	On-target	10	0.1636	0.1631	0.1604	0.1587	0.1579	0.0516	0.0509	0.0505	0.0498	0.0494	0.0234	0.0231	0.0229	0.0227	0.0224
Chisq(5)	On-target	15	0.1607	0.1613	0.1590	0.1583	0.1564	0.0511	0.0504	0.0500	0.0495	0.0491	0.0232	0.0229	0.0228	0.0226	0.0222
Chisq(5)	On-target	20	0.1616	0.1609	0.1592	0.1577	0.1566	0.0506	0.0504	0.0499	0.0495	0.0490	0.0231	0.0228	0.0227	0.0224	0.0223
Chisq(5)	Off by 0.5	3	0.1742	0.1703	0.1652	0.1633	0.1610	0.0551	0.0529	0.0519	0.0513	0.0510	0.0248	0.0242	0.0238	0.0234	0.0230
Chisq(5)	Off by 0.5	5	0.1694	0.1659	0.1624	0.1602	0.1581	0.0529	0.0525	0.0510	0.0502	0.0501	0.0241	0.0238	0.0233	0.0229	0.0226
Chisq(5)	Off by 0.5	10	0.1662	0.1640	0.1593	0.1584	0.1566	0.0524	0.0516	0.0503	0.0496	0.0491	0.0236	0.0233	0.0227	0.0225	0.0222
Chisq(5)	Off by 0.5	15	0.1646	0.1614	0.1591	0.1581	0.1564	0.0519	0.0508	0.0500	0.0494	0.0491	0.0237	0.0231	0.0228	0.0223	0.0222
Chisq(5)	Off by 0.5	20	0.1639	0.1618	0.1583	0.1575	0.1556	0.0517	0.0509	0.0500	0.0494	0.0487	0.0235	0.0231	0.0227	0.0224	0.0222
Chisq(5)	Off by 1.0	3	0.1744	0.1714	0.1665	0.1636	0.1617	0.0553	0.0538	0.0519	0.0514	0.0507	0.0246	0.0244	0.0237	0.0235	0.0230
Chisq(5)	Off by 1.0	5	0.1699	0.1670	0.1632	0.1612	0.1581	0.0543	0.0527	0.0514	0.0502	0.0499	0.0243	0.0240	0.0231	0.0228	0.0226
Chisq(5)	Off by 1.0	10	0.1671	0.1638	0.1602	0.1585	0.1563	0.0527	0.0517	0.0504	0.0498	0.0490	0.0242	0.0234	0.0229	0.0225	0.0223
Chisq(5)	Off by 1.0	15	0.1669	0.1627	0.1591	0.1576	0.1551	0.0528	0.0516	0.0500	0.0496	0.0490	0.0238	0.0234	0.0228	0.0225	0.0221
Chisq(5)	Off by 1.0	20	0.1662	0.1625	0.1586	0.1567	0.1555	0.0527	0.0512	0.0501	0.0494	0.0488	0.0237	0.0232	0.0227	0.0223	0.0221
Chisq(5)	Off by 2.5	3						0.0592	0.0569	0.0529	0.0513	0.0505	0.0270	0.0254	0.0241	0.0239	0.0228
Chisq(5)	Off by 2.5	5						0.0583	0.0548	0.0521	0.0504	0.0500	0.0266	0.0247	0.0235	0.0233	0.0227
Chisq(5)	Off by 2.5	10						0.0565	0.0535	0.0513	0.0501	0.0490	0.0257	0.0244	0.0232	0.0227	0.0222
Chisq(5)	Off by 2.5	15						0.0563	0.0532	0.0509	0.0496	0.0490	0.0255	0.0243	0.0231	0.0226	0.0222
Chisq(5)	Off by 2.5	20						0.0556	0.0532	0.0506	0.0495	0.0489	0.0255	0.0241	0.0230	0.0225	0.0221

Table 23 Standard Error Estimated for 95th Percentile by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.2186	0.2055	0.1947	0.1891	0.1864	0.0674	0.0647	0.0615	0.0595	0.0579	0.0313	0.0299	0.0280	0.0275	0.0265
Normal	On-target	5	0.2126	0.2036	0.1921	0.1861	0.1818	0.0670	0.0639	0.0605	0.0589	0.0576	0.0302	0.0291	0.0278	0.0267	0.0261
Normal	On-target	10	0.2073	0.1980	0.1899	0.1839	0.1801	0.0652	0.0626	0.0594	0.0577	0.0567	0.0299	0.0283	0.0271	0.0263	0.0258
Normal	On-target	15	0.2044	0.1969	0.1885	0.1827	0.1790	0.0645	0.0617	0.0591	0.0574	0.0565	0.0294	0.0283	0.0268	0.0260	0.0256
Normal	On-target	20	0.2038	0.1963	0.1881	0.1825	0.1785	0.0646	0.0617	0.0590	0.0574	0.0563	0.0294	0.0281	0.0268	0.0261	0.0256
Normal	Off by 0.5	3	0.2143	0.2048	0.1962	0.1882	0.1839	0.0694	0.0652	0.0612	0.0592	0.0583	0.0307	0.0294	0.0278	0.0270	0.0265
Normal	Off by 0.5	5	0.2103	0.2009	0.1909	0.1846	0.1815	0.0669	0.0631	0.0604	0.0583	0.0572	0.0302	0.0290	0.0275	0.0264	0.0260
Normal	Off by 0.5	10	0.2059	0.1982	0.1882	0.1835	0.1796	0.0651	0.0624	0.0590	0.0572	0.0564	0.0295	0.0281	0.0270	0.0261	0.0256
Normal	Off by 0.5	15	0.2048	0.1965	0.1876	0.1824	0.1794	0.0644	0.0616	0.0589	0.0571	0.0561	0.0292	0.0281	0.0266	0.0260	0.0254
Normal	Off by 0.5	20	0.2036	0.1955	0.1869	0.1815	0.1792	0.0642	0.0614	0.0588	0.0568	0.0559	0.0292	0.0281	0.0266	0.0259	0.0253
Normal	Off by 1.0	3	0.2181	0.2047	0.1965	0.1874	0.1839	0.0676	0.0645	0.0606	0.0598	0.0583	0.0311	0.0296	0.0278	0.0266	0.0263
Normal	Off by 1.0	5	0.2088	0.2011	0.1909	0.1846	0.1810	0.0657	0.0635	0.0598	0.0577	0.0571	0.0297	0.0288	0.0271	0.0264	0.0258
Normal	Off by 1.0	10	0.2042	0.1956	0.1887	0.1824	0.1796	0.0650	0.0621	0.0590	0.0576	0.0561	0.0295	0.0283	0.0268	0.0261	0.0255
Normal	Off by 1.0	15	0.2043	0.1957	0.1871	0.1818	0.1790	0.0642	0.0616	0.0588	0.0569	0.0560	0.0293	0.0280	0.0265	0.0259	0.0253
Normal	Off by 1.0	20	0.2039	0.1954	0.1870	0.1811	0.1785	0.0641	0.0613	0.0585	0.0569	0.0560	0.0291	0.0280	0.0265	0.0258	0.0253
Normal	Off by 2.5	3						0.0674	0.0651	0.0618	0.0596	0.0575	0.0310	0.0295	0.0278	0.0269	0.0265
Normal	Off by 2.5	5						0.0659	0.0637	0.0605	0.0582	0.0569	0.0301	0.0289	0.0273	0.0263	0.0261
Normal	Off by 2.5	10						0.0649	0.0621	0.0592	0.0576	0.0562	0.0297	0.0282	0.0269	0.0261	0.0256
Normal	Off by 2.5	15						0.0645	0.0615	0.0587	0.0571	0.0559	0.0293	0.0280	0.0269	0.0259	0.0254
Normal	Off by 2.5	20						0.0645	0.0612	0.0585	0.0569	0.0558	0.0292	0.0280	0.0268	0.0258	0.0253
Uniform	On-target	3	0.2167	0.2024	0.1808	0.1686	0.1634	0.0685	0.0630	0.0575	0.0533	0.0514	0.0313	0.0288	0.0264	0.0246	0.0236
Uniform	On-target	5	0.2153	0.1964	0.1778	0.1673	0.1607	0.0670	0.0623	0.0562	0.0530	0.0510	0.0304	0.0279	0.0257	0.0239	0.0230
Uniform	On-target	10	0.2083	0.1937	0.1746	0.1644	0.1584	0.0655	0.0609	0.0553	0.0521	0.0501	0.0298	0.0279	0.0253	0.0239	0.0228
Uniform	On-target	15	0.2075	0.1923	0.1743	0.1640	0.1580	0.0650	0.0607	0.0549	0.0520	0.0500	0.0298	0.0275	0.0252	0.0237	0.0228
Uniform	On-target	20	0.2067	0.1913	0.1733	0.1643	0.1579	0.0649	0.0605	0.0549	0.0519	0.0499	0.0294	0.0276	0.0251	0.0236	0.0227
Uniform	Off by 0.5	3	0.2160	0.1986	0.1778	0.1664	0.1606	0.0687	0.0629	0.0561	0.0531	0.0516	0.0305	0.0291	0.0260	0.0242	0.0232
Uniform	Off by 0.5	5	0.2104	0.1944	0.1755	0.1657	0.1588	0.0664	0.0616	0.0556	0.0521	0.0497	0.0303	0.0278	0.0254	0.0237	0.0229
Uniform	Off by 0.5	10	0.2062	0.1915	0.1733	0.1631	0.1570	0.0652	0.0601	0.0545	0.0516	0.0497	0.0298	0.0273	0.0249	0.0234	0.0226
Uniform	Off by 0.5	15	0.2065	0.1901	0.1718	0.1628	0.1563	0.0650	0.0601	0.0544	0.0512	0.0495	0.0296	0.0273	0.0248	0.0234	0.0225
Uniform	Off by 0.5	20	0.2055	0.1897	0.1714	0.1624	0.1561	0.0644	0.0597	0.0541	0.0510	0.0494	0.0295	0.0272	0.0247	0.0233	0.0224
Uniform	Off by 1.0	3	0.2151	0.1974	0.1761	0.1657	0.1589	0.0681	0.0623	0.0563	0.0526	0.0506	0.0311	0.0280	0.0257	0.0240	0.0229
Uniform	Off by 1.0	5	0.2113	0.1930	0.1738	0.1637	0.1570	0.0663	0.0610	0.0545	0.0518	0.0500	0.0301	0.0279	0.0253	0.0236	0.0227
Uniform	Off by 1.0	10	0.2058	0.1909	0.1711	0.1617	0.1566	0.0649	0.0599	0.0544	0.0512	0.0493	0.0293	0.0272	0.0247	0.0233	0.0226
Uniform	Off by 1.0	15	0.2053	0.1888	0.1709	0.1609	0.1556	0.0644	0.0595	0.0540	0.0510	0.0491	0.0294	0.0270	0.0246	0.0232	0.0224
Uniform	Off by 1.0	20	0.2048	0.1882	0.1704	0.1610	0.1555	0.0642	0.0595	0.0539	0.0510	0.0491	0.0293	0.0272	0.0246	0.0231	0.0223
Uniform	Off by 2.5	3						0.0676	0.0628	0.0554	0.0523	0.0500	0.0312	0.0281	0.0254	0.0240	0.0227
Uniform	Off by 2.5	5						0.0667	0.0604	0.0549	0.0515	0.0496	0.0303	0.0279	0.0249	0.0235	0.0224
Uniform	Off by 2.5	10						0.0643	0.0594	0.0539	0.0509	0.0490	0.0293	0.0274	0.0246	0.0231	0.0222
Uniform	Off by 2.5	15						0.0639	0.0593	0.0535	0.0507	0.0489	0.0292	0.0270	0.0246	0.0230	0.0223
Uniform	Off by 2.5	20						0.0635	0.0591	0.0533	0.0506	0.0487	0.0291	0.0271	0.0244	0.0230	0.0221
Bimodal	On-target	3	0.2142	0.1986	0.1750	0.1631	0.1567	0.0688	0.0631	0.0549	0.0516	0.0492	0.0313	0.0289	0.0257	0.0236	0.0224

Appendix 4

Standard Error Estimates of Simulation

Bimodal	On-target	5	0.2126	0.1942	0.1727	0.1617	0.1545	0.0667	0.0615	0.0542	0.0510	0.0483	0.0304	0.0282	0.0249	0.0232	0.0222
Bimodal	On-target	10	0.2065	0.1913	0.1698	0.1596	0.1517	0.0650	0.0604	0.0536	0.0503	0.0480	0.0297	0.0275	0.0246	0.0229	0.0220
Bimodal	On-target	15	0.2062	0.1900	0.1706	0.1584	0.1513	0.0650	0.0601	0.0534	0.0500	0.0479	0.0297	0.0272	0.0244	0.0228	0.0219
Bimodal	On-target	20	0.2043	0.1896	0.1698	0.1584	0.1508	0.0647	0.0597	0.0534	0.0498	0.0479	0.0294	0.0273	0.0244	0.0227	0.0218
Bimodal	Off by 0.5	3	0.2141	0.1941	0.1735	0.1629	0.1547	0.0668	0.0624	0.0543	0.0513	0.0482	0.0308	0.0287	0.0254	0.0235	0.0225
Bimodal	Off by 0.5	5	0.2097	0.1905	0.1724	0.1597	0.1524	0.0655	0.0608	0.0541	0.0501	0.0485	0.0305	0.0277	0.0244	0.0230	0.0220
Bimodal	Off by 0.5	10	0.2043	0.1879	0.1687	0.1576	0.1505	0.0648	0.0597	0.0532	0.0498	0.0474	0.0295	0.0273	0.0243	0.0226	0.0217
Bimodal	Off by 0.5	15	0.2048	0.1872	0.1680	0.1566	0.1495	0.0644	0.0594	0.0529	0.0495	0.0474	0.0295	0.0270	0.0240	0.0225	0.0217
Bimodal	Off by 0.5	20	0.2037	0.1872	0.1681	0.1566	0.1492	0.0641	0.0591	0.0528	0.0492	0.0474	0.0292	0.0269	0.0241	0.0224	0.0215
Bimodal	Off by 1.0	3	0.2149	0.1964	0.1732	0.1629	0.1522	0.0678	0.0617	0.0547	0.0503	0.0488	0.0311	0.0284	0.0245	0.0231	0.0223
Bimodal	Off by 1.0	5	0.2081	0.1904	0.1707	0.1583	0.1510	0.0659	0.0604	0.0537	0.0501	0.0482	0.0302	0.0276	0.0243	0.0229	0.0219
Bimodal	Off by 1.0	10	0.2041	0.1876	0.1682	0.1578	0.1498	0.0646	0.0590	0.0528	0.0492	0.0474	0.0299	0.0270	0.0241	0.0227	0.0216
Bimodal	Off by 1.0	15	0.2038	0.1863	0.1666	0.1560	0.1494	0.0645	0.0588	0.0526	0.0492	0.0472	0.0293	0.0269	0.0239	0.0225	0.0215
Bimodal	Off by 1.0	20	0.2026	0.1863	0.1661	0.1559	0.1488	0.0642	0.0588	0.0524	0.0490	0.0470	0.0292	0.0268	0.0238	0.0223	0.0215
Bimodal	Off by 2.5	3						0.0686	0.0620	0.0550	0.0503	0.0484	0.0305	0.0283	0.0250	0.0231	0.0224
Bimodal	Off by 2.5	5						0.0663	0.0605	0.0533	0.0498	0.0478	0.0301	0.0275	0.0244	0.0231	0.0218
Bimodal	Off by 2.5	10						0.0644	0.0586	0.0528	0.0492	0.0471	0.0293	0.0274	0.0241	0.0224	0.0215
Bimodal	Off by 2.5	15						0.0633	0.0586	0.0523	0.0491	0.0469	0.0290	0.0269	0.0240	0.0223	0.0214
Bimodal	Off by 2.5	20						0.0634	0.0583	0.0521	0.0490	0.0467	0.0290	0.0268	0.0239	0.0223	0.0214
Chisq(5)	On-target	3	0.2285	0.2444	0.2612	0.2676	0.2763	0.0703	0.0739	0.0804	0.0828	0.0852	0.0321	0.0344	0.0366	0.0381	0.0383
Chisq(5)	On-target	5	0.2224	0.2375	0.2553	0.2684	0.2758	0.0693	0.0737	0.0782	0.0817	0.0838	0.0314	0.0332	0.0363	0.0374	0.0381
Chisq(5)	On-target	10	0.2191	0.2333	0.2527	0.2654	0.2733	0.0682	0.0723	0.0779	0.0805	0.0830	0.0309	0.0329	0.0356	0.0370	0.0377
Chisq(5)	On-target	15	0.2162	0.2321	0.2518	0.2650	0.2724	0.0678	0.0715	0.0775	0.0803	0.0824	0.0305	0.0327	0.0354	0.0368	0.0375
Chisq(5)	On-target	20	0.2156	0.2317	0.2522	0.2644	0.2723	0.0672	0.0716	0.0773	0.0801	0.0824	0.0305	0.0326	0.0352	0.0366	0.0373
Chisq(5)	Off by 0.5	3	0.2344	0.2455	0.2558	0.2675	0.2769	0.0729	0.0759	0.0802	0.0827	0.0836	0.0335	0.0345	0.0366	0.0379	0.0385
Chisq(5)	Off by 0.5	5	0.2251	0.2406	0.2551	0.2672	0.2728	0.0711	0.0751	0.0796	0.0814	0.0830	0.0324	0.0340	0.0360	0.0371	0.0380
Chisq(5)	Off by 0.5	10	0.2217	0.2353	0.2519	0.2650	0.2700	0.0695	0.0733	0.0779	0.0806	0.0824	0.0315	0.0333	0.0356	0.0367	0.0375
Chisq(5)	Off by 0.5	15	0.2205	0.2343	0.2517	0.2636	0.2704	0.0691	0.0728	0.0777	0.0802	0.0818	0.0311	0.0330	0.0355	0.0368	0.0372
Chisq(5)	Off by 0.5	20	0.2190	0.2338	0.2517	0.2644	0.2696	0.0686	0.0725	0.0773	0.0801	0.0819	0.0310	0.0330	0.0353	0.0366	0.0372
Chisq(5)	Off by 1.0	3	0.2366	0.2460	0.2596	0.2695	0.2729	0.0729	0.0768	0.0801	0.0817	0.0838	0.0333	0.0355	0.0369	0.0373	0.0379
Chisq(5)	Off by 1.0	5	0.2306	0.2420	0.2546	0.2670	0.2712	0.0724	0.0752	0.0790	0.0809	0.0830	0.0324	0.0340	0.0362	0.0369	0.0376
Chisq(5)	Off by 1.0	10	0.2252	0.2374	0.2532	0.2632	0.2695	0.0708	0.0736	0.0782	0.0800	0.0822	0.0320	0.0336	0.0356	0.0365	0.0372
Chisq(5)	Off by 1.0	15	0.2246	0.2350	0.2517	0.2635	0.2687	0.0699	0.0740	0.0776	0.0797	0.0815	0.0317	0.0333	0.0355	0.0364	0.0370
Chisq(5)	Off by 1.0	20	0.2227	0.2354	0.2518	0.2622	0.2685	0.0700	0.0735	0.0773	0.0797	0.0817	0.0316	0.0331	0.0353	0.0363	0.0369
Chisq(5)	Off by 2.5	3						0.0786	0.0765	0.0785	0.0795	0.0798	0.0357	0.0347	0.0356	0.0363	0.0372
Chisq(5)	Off by 2.5	5						0.0758	0.0763	0.0773	0.0785	0.0799	0.0350	0.0342	0.0349	0.0358	0.0361
Chisq(5)	Off by 2.5	10						0.0749	0.0743	0.0761	0.0776	0.0790	0.0341	0.0338	0.0346	0.0354	0.0358
Chisq(5)	Off by 2.5	15						0.0743	0.0737	0.0756	0.0772	0.0784	0.0338	0.0335	0.0342	0.0352	0.0357
Chisq(5)	Off by 2.5	20						0.0736	0.0740	0.0755	0.0772	0.0786	0.0336	0.0334	0.0342	0.0351	0.0357

Table 24 Standard Error Estimated for 97.5th Percentile by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Number of PVs	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	3	0.3002	0.2854	0.2774	0.2683	0.2645	0.0924	0.0878	0.0850	0.0813	0.0802	0.0418	0.0402	0.0383	0.0380	0.0374
Normal	On-target	5	0.2942	0.2828	0.2723	0.2649	0.2604	0.0900	0.0867	0.0829	0.0816	0.0792	0.0408	0.0399	0.0374	0.0365	0.0362
Normal	On-target	10	0.2874	0.2793	0.2712	0.2614	0.2576	0.0879	0.0852	0.0816	0.0796	0.0776	0.0398	0.0386	0.0370	0.0364	0.0356
Normal	On-target	15	0.2852	0.2768	0.2676	0.2600	0.2551	0.0877	0.0848	0.0810	0.0793	0.0772	0.0397	0.0385	0.0369	0.0361	0.0354
Normal	On-target	20	0.2851	0.2762	0.2680	0.2608	0.2563	0.0874	0.0839	0.0811	0.0791	0.0773	0.0397	0.0383	0.0367	0.0360	0.0355
Normal	Off by 0.5	3	0.2948	0.2889	0.2764	0.2644	0.2619	0.0925	0.0881	0.0832	0.0805	0.0797	0.0422	0.0399	0.0380	0.0369	0.0369
Normal	Off by 0.5	5	0.2949	0.2798	0.2717	0.2609	0.2601	0.0900	0.0855	0.0821	0.0796	0.0788	0.0411	0.0392	0.0377	0.0365	0.0357
Normal	Off by 0.5	10	0.2854	0.2765	0.2662	0.2604	0.2560	0.0880	0.0844	0.0810	0.0784	0.0775	0.0401	0.0384	0.0369	0.0359	0.0353
Normal	Off by 0.5	15	0.2849	0.2760	0.2669	0.2602	0.2567	0.0869	0.0836	0.0805	0.0780	0.0770	0.0400	0.0383	0.0367	0.0358	0.0352
Normal	Off by 0.5	20	0.2845	0.2759	0.2659	0.2593	0.2574	0.0871	0.0837	0.0801	0.0780	0.0770	0.0398	0.0381	0.0365	0.0357	0.0351
Normal	Off by 1.0	3	0.3047	0.2919	0.2711	0.2638	0.2621	0.0912	0.0873	0.0826	0.0809	0.0792	0.0414	0.0403	0.0386	0.0370	0.0365
Normal	Off by 1.0	5	0.2957	0.2796	0.2702	0.2608	0.2573	0.0890	0.0861	0.0808	0.0798	0.0783	0.0411	0.0394	0.0373	0.0362	0.0359
Normal	Off by 1.0	10	0.2863	0.2753	0.2686	0.2607	0.2575	0.0875	0.0842	0.0805	0.0786	0.0774	0.0398	0.0383	0.0369	0.0358	0.0353
Normal	Off by 1.0	15	0.2838	0.2759	0.2652	0.2595	0.2553	0.0871	0.0836	0.0797	0.0781	0.0771	0.0395	0.0380	0.0368	0.0357	0.0351
Normal	Off by 1.0	20	0.2840	0.2752	0.2662	0.2593	0.2559	0.0870	0.0835	0.0799	0.0779	0.0770	0.0393	0.0381	0.0364	0.0355	0.0349
Normal	Off by 2.5	3						0.0916	0.0871	0.0832	0.0802	0.0787	0.0416	0.0399	0.0383	0.0368	0.0357
Normal	Off by 2.5	5						0.0891	0.0855	0.0816	0.0793	0.0780	0.0403	0.0392	0.0376	0.0364	0.0358
Normal	Off by 2.5	10						0.0877	0.0841	0.0807	0.0785	0.0770	0.0398	0.0383	0.0367	0.0357	0.0351
Normal	Off by 2.5	15						0.0868	0.0832	0.0803	0.0778	0.0764	0.0395	0.0380	0.0367	0.0355	0.0350
Normal	Off by 2.5	20						0.0868	0.0835	0.0798	0.0777	0.0765	0.0396	0.0377	0.0364	0.0355	0.0349
Uniform	On-target	3	0.3001	0.2621	0.2255	0.2000	0.1870	0.0918	0.0818	0.0700	0.0634	0.0595	0.0409	0.0373	0.0324	0.0293	0.0274
Uniform	On-target	5	0.2900	0.2575	0.2203	0.1981	0.1844	0.0884	0.0801	0.0686	0.0633	0.0587	0.0396	0.0369	0.0313	0.0293	0.0270
Uniform	On-target	10	0.2842	0.2550	0.2157	0.1944	0.1816	0.0871	0.0787	0.0674	0.0616	0.0577	0.0393	0.0361	0.0311	0.0285	0.0266
Uniform	On-target	15	0.2829	0.2526	0.2149	0.1938	0.1811	0.0864	0.0779	0.0674	0.0616	0.0575	0.0393	0.0357	0.0310	0.0283	0.0264
Uniform	On-target	20	0.2818	0.2503	0.2139	0.1934	0.1809	0.0860	0.0775	0.0671	0.0614	0.0574	0.0391	0.0355	0.0309	0.0283	0.0263
Uniform	Off by 0.5	3	0.2937	0.2557	0.2147	0.1955	0.1835	0.0900	0.0807	0.0684	0.0626	0.0591	0.0404	0.0371	0.0319	0.0290	0.0266
Uniform	Off by 0.5	5	0.2854	0.2518	0.2143	0.1943	0.1796	0.0877	0.0790	0.0679	0.0613	0.0573	0.0403	0.0360	0.0310	0.0284	0.0263
Uniform	Off by 0.5	10	0.2825	0.2481	0.2111	0.1909	0.1784	0.0864	0.0769	0.0664	0.0606	0.0564	0.0394	0.0355	0.0301	0.0280	0.0257
Uniform	Off by 0.5	15	0.2791	0.2470	0.2096	0.1909	0.1785	0.0859	0.0763	0.0656	0.0601	0.0567	0.0388	0.0352	0.0301	0.0276	0.0256
Uniform	Off by 0.5	20	0.2807	0.2479	0.2092	0.1897	0.1771	0.0851	0.0764	0.0657	0.0600	0.0565	0.0388	0.0349	0.0300	0.0277	0.0256
Uniform	Off by 1.0	3	0.2858	0.2554	0.2150	0.1934	0.1810	0.0894	0.0790	0.0679	0.0623	0.0579	0.0404	0.0371	0.0310	0.0282	0.0262
Uniform	Off by 1.0	5	0.2851	0.2465	0.2123	0.1904	0.1780	0.0876	0.0776	0.0661	0.0600	0.0568	0.0398	0.0357	0.0305	0.0278	0.0259
Uniform	Off by 1.0	10	0.2772	0.2446	0.2071	0.1876	0.1774	0.0856	0.0757	0.0654	0.0596	0.0562	0.0389	0.0351	0.0301	0.0275	0.0254
Uniform	Off by 1.0	15	0.2766	0.2433	0.2071	0.1877	0.1765	0.0847	0.0754	0.0649	0.0593	0.0559	0.0385	0.0347	0.0298	0.0272	0.0254
Uniform	Off by 1.0	20	0.2760	0.2434	0.2061	0.1867	0.1755	0.0848	0.0754	0.0647	0.0592	0.0556	0.0386	0.0347	0.0298	0.0272	0.0253
Uniform	Off by 2.5	3						0.0893	0.0793	0.0666	0.0600	0.0567	0.0400	0.0359	0.0307	0.0280	0.0260
Uniform	Off by 2.5	5						0.0867	0.0761	0.0651	0.0595	0.0560	0.0393	0.0351	0.0303	0.0273	0.0253
Uniform	Off by 2.5	10						0.0839	0.0749	0.0646	0.0591	0.0554	0.0387	0.0349	0.0300	0.0268	0.0252
Uniform	Off by 2.5	15						0.0840	0.0751	0.0644	0.0587	0.0552	0.0384	0.0346	0.0298	0.0266	0.0250
Uniform	Off by 2.5	20						0.0836	0.0744	0.0639	0.0584	0.0550	0.0383	0.0345	0.0295	0.0267	0.0249
Bimodal	On-target	3	0.2924	0.2583	0.2159	0.1959	0.1827	0.0915	0.0803	0.0688	0.0629	0.0585	0.0418	0.0380	0.0319	0.0284	0.0264

Appendix 4

Standard Error Estimates of Simulation

Bimodal	On-target	5	0.2882	0.2567	0.2149	0.1944	0.1810	0.0879	0.0793	0.0671	0.0612	0.0573	0.0404	0.0363	0.0310	0.0279	0.0262
Bimodal	On-target	10	0.2814	0.2496	0.2113	0.1897	0.1780	0.0861	0.0775	0.0663	0.0603	0.0562	0.0391	0.0356	0.0305	0.0276	0.0258
Bimodal	On-target	15	0.2818	0.2477	0.2100	0.1896	0.1766	0.0856	0.0771	0.0654	0.0600	0.0560	0.0392	0.0353	0.0303	0.0274	0.0255
Bimodal	On-target	20	0.2810	0.2478	0.2107	0.1893	0.1755	0.0862	0.0766	0.0652	0.0597	0.0560	0.0391	0.0352	0.0304	0.0274	0.0256
Bimodal	Off by 0.5	3	0.2900	0.2531	0.2134	0.1937	0.1809	0.0893	0.0796	0.0664	0.0613	0.0569	0.0405	0.0375	0.0310	0.0279	0.0263
Bimodal	Off by 0.5	5	0.2845	0.2495	0.2115	0.1897	0.1773	0.0864	0.0771	0.0662	0.0602	0.0565	0.0396	0.0356	0.0305	0.0277	0.0262
Bimodal	Off by 0.5	10	0.2790	0.2430	0.2082	0.1879	0.1748	0.0854	0.0755	0.0648	0.0589	0.0551	0.0389	0.0349	0.0301	0.0272	0.0254
Bimodal	Off by 0.5	15	0.2782	0.2431	0.2063	0.1862	0.1738	0.0848	0.0756	0.0647	0.0587	0.0552	0.0387	0.0348	0.0298	0.0269	0.0254
Bimodal	Off by 0.5	20	0.2754	0.2423	0.2054	0.1860	0.1741	0.0843	0.0751	0.0644	0.0586	0.0552	0.0385	0.0346	0.0297	0.0268	0.0251
Bimodal	Off by 1.0	3	0.2955	0.2522	0.2139	0.1919	0.1798	0.0886	0.0781	0.0665	0.0603	0.0565	0.0406	0.0368	0.0302	0.0282	0.0261
Bimodal	Off by 1.0	5	0.2750	0.2487	0.2059	0.1890	0.1750	0.0865	0.0759	0.0650	0.0595	0.0559	0.0399	0.0354	0.0302	0.0274	0.0255
Bimodal	Off by 1.0	10	0.2762	0.2411	0.2051	0.1864	0.1736	0.0851	0.0749	0.0639	0.0583	0.0549	0.0391	0.0345	0.0297	0.0268	0.0250
Bimodal	Off by 1.0	15	0.2766	0.2405	0.2029	0.1847	0.1723	0.0846	0.0743	0.0638	0.0580	0.0545	0.0384	0.0341	0.0296	0.0267	0.0250
Bimodal	Off by 1.0	20	0.2740	0.2395	0.2023	0.1847	0.1723	0.0842	0.0744	0.0636	0.0580	0.0543	0.0384	0.0343	0.0294	0.0265	0.0249
Bimodal	Off by 2.5	3						0.0886	0.0776	0.0658	0.0609	0.0562	0.0397	0.0364	0.0304	0.0270	0.0256
Bimodal	Off by 2.5	5						0.0867	0.0762	0.0641	0.0589	0.0551	0.0393	0.0346	0.0296	0.0270	0.0252
Bimodal	Off by 2.5	10						0.0845	0.0739	0.0635	0.0579	0.0542	0.0384	0.0343	0.0291	0.0263	0.0246
Bimodal	Off by 2.5	15						0.0836	0.0739	0.0631	0.0576	0.0539	0.0381	0.0341	0.0288	0.0262	0.0246
Bimodal	Off by 2.5	20						0.0834	0.0733	0.0629	0.0573	0.0537	0.0381	0.0338	0.0287	0.0262	0.0246
Chisq(5)	On-target	3	0.3401	0.4123	0.5657	0.6558	0.7110	0.1009	0.1194	0.1504	0.1702	0.1859	0.0465	0.0554	0.0674	0.0761	0.0826
Chisq(5)	On-target	5	0.3332	0.4095	0.5485	0.6433	0.7181	0.0992	0.1185	0.1494	0.1698	0.1857	0.0456	0.0542	0.0673	0.0761	0.0821
Chisq(5)	On-target	10	0.3292	0.4079	0.5372	0.6445	0.7222	0.0976	0.1167	0.1483	0.1685	0.1847	0.0446	0.0535	0.0663	0.0753	0.0817
Chisq(5)	On-target	15	0.3275	0.4062	0.5419	0.6408	0.7111	0.0971	0.1161	0.1479	0.1683	0.1842	0.0443	0.0533	0.0663	0.0752	0.0817
Chisq(5)	On-target	20	0.3265	0.4039	0.5374	0.6414	0.7192	0.0968	0.1155	0.1480	0.1680	0.1842	0.0442	0.0532	0.0659	0.0751	0.0815
Chisq(5)	Off by 0.5	3	0.3487	0.4307	0.5590	0.6770	0.7390	0.1031	0.1253	0.1543	0.1743	0.1865	0.0470	0.0563	0.0692	0.0773	0.0842
Chisq(5)	Off by 0.5	5	0.3447	0.4348	0.5671	0.6792	0.7332	0.1024	0.1219	0.1525	0.1739	0.1869	0.0473	0.0557	0.0689	0.0772	0.0829
Chisq(5)	Off by 0.5	10	0.3389	0.4216	0.5624	0.6804	0.7483	0.1000	0.1211	0.1511	0.1719	0.1850	0.0460	0.0552	0.0680	0.0766	0.0825
Chisq(5)	Off by 0.5	15	0.3415	0.4222	0.5616	0.6669	0.7354	0.1002	0.1210	0.1512	0.1720	0.1855	0.0458	0.0547	0.0677	0.0764	0.0824
Chisq(5)	Off by 0.5	20	0.3367	0.4210	0.5642	0.6727	0.7364	0.0995	0.1200	0.1508	0.1720	0.1856	0.0456	0.0545	0.0676	0.0765	0.0825
Chisq(5)	Off by 1.0	3	0.3627	0.4457	0.5686	0.7037	0.7299	0.1060	0.1254	0.1544	0.1730	0.1878	0.0484	0.0582	0.0693	0.0775	0.0833
Chisq(5)	Off by 1.0	5	0.3548	0.4423	0.5717	0.6953	0.7346	0.1040	0.1244	0.1538	0.1719	0.1877	0.0478	0.0568	0.0690	0.0768	0.0826
Chisq(5)	Off by 1.0	10	0.3492	0.4345	0.5778	0.6816	0.7455	0.1031	0.1230	0.1528	0.1715	0.1872	0.0471	0.0557	0.0685	0.0765	0.0822
Chisq(5)	Off by 1.0	15	0.3495	0.4342	0.5703	0.6727	0.7505	0.1017	0.1230	0.1519	0.1711	0.1866	0.0466	0.0556	0.0681	0.0765	0.0822
Chisq(5)	Off by 1.0	20	0.3475	0.4325	0.5709	0.6982	0.7493	0.1022	0.1218	0.1520	0.1711	0.1863	0.0467	0.0554	0.0681	0.0762	0.0819
Chisq(5)	Off by 2.5	3						0.1138	0.1226	0.1454	0.1600	0.1722	0.0510	0.0564	0.0656	0.0715	0.0770
Chisq(5)	Off by 2.5	5						0.1096	0.1214	0.1441	0.1578	0.1720	0.0504	0.0553	0.0648	0.0706	0.0760
Chisq(5)	Off by 2.5	10						0.1088	0.1200	0.1423	0.1575	0.1706	0.0496	0.0545	0.0638	0.0704	0.0754
Chisq(5)	Off by 2.5	15						0.1078	0.1196	0.1418	0.1573	0.1700	0.0493	0.0545	0.0637	0.0703	0.0754
Chisq(5)	Off by 2.5	20						0.1073	0.1193	0.1416	0.1568	0.1702	0.0492	0.0543	0.0636	0.0701	0.0751

Appendix 5 Estimated Bias and RMSEs of Simulation Two

Table 25 Bias and RMSE of the Generating Regression Coefficients

Target	Number of PVs	Bias				RMSE			
		10	20	40	60	10	20	40	60
$\beta_1 = -0.09$ (Gender)									
On-target	1	0.0000	0.0000	0.0000	0.0001	0.0059	0.0051	0.0046	0.0047
	3	0.0000	0.0000	0.0001	0.0000	0.0056	0.0048	0.0045	0.0044
	5	0.0000	0.0000	0.0001	0.0000	0.0055	0.0048	0.0044	0.0045
	10	0.0000	0.0000	0.0001	0.0000	0.0056	0.0047	0.0043	0.0044
	15	0.0000	0.0000	0.0001	0.0000	0.0055	0.0047	0.0044	0.0044
	20	-0.0001	0.0000	0.0001	0.0000	0.0055	0.0047	0.0044	0.0044
Off by 0.5	1	-0.0004	-0.0001	-0.0001	-0.0001	0.0063	0.0051	0.0044	0.0043
	3	-0.0004	-0.0002	-0.0001	-0.0001	0.0059	0.0049	0.0043	0.0042
	5	-0.0003	-0.0002	-0.0001	-0.0001	0.0059	0.0048	0.0043	0.0042
	10	-0.0002	-0.0002	-0.0001	-0.0001	0.0058	0.0047	0.0043	0.0042
	15	-0.0003	-0.0002	-0.0001	-0.0001	0.0058	0.0047	0.0042	0.0042
	20	-0.0002	-0.0002	-0.0001	-0.0001	0.0058	0.0047	0.0042	0.0041
Off by 1.0	1	-0.0003	-0.0002	0.0000	0.0000	0.0060	0.0054	0.0047	0.0044
	3	-0.0002	-0.0001	0.0000	-0.0001	0.0058	0.0050	0.0045	0.0042
	5	-0.0001	-0.0001	0.0001	0.0000	0.0057	0.0049	0.0045	0.0043
	10	-0.0001	-0.0001	0.0000	-0.0001	0.0056	0.0049	0.0045	0.0042
	15	-0.0001	-0.0001	0.0000	-0.0001	0.0056	0.0049	0.0044	0.0042
	20	-0.0001	-0.0001	0.0000	-0.0001	0.0056	0.0049	0.0044	0.0042
$\beta_2 = -0.57$ (LBOTE)									
On-target	1	-0.0021	-0.0014	-0.0016	-0.0022	0.0684	0.0572	0.0529	0.0502
	3	-0.0005	-0.0019	-0.0011	-0.0025	0.0629	0.0534	0.0505	0.0492
	5	-0.0001	-0.0010	-0.0005	-0.0026	0.0609	0.0525	0.0499	0.0488
	10	-0.0005	-0.0013	-0.0007	-0.0026	0.0618	0.0516	0.0498	0.0484
	15	-0.0006	-0.0019	-0.0006	-0.0025	0.0612	0.0517	0.0494	0.0484
	20	-0.0006	-0.0017	-0.0007	-0.0025	0.0610	0.0516	0.0492	0.0483
Off by 0.5	1	-0.0046	-0.0001	0.0016	-0.0007	0.0685	0.0583	0.0504	0.0522
	3	-0.0054	-0.0013	0.0010	-0.0009	0.0646	0.0539	0.0484	0.0501
	5	-0.0055	-0.0012	0.0017	-0.0006	0.0638	0.0534	0.0479	0.0496
	10	-0.0048	-0.0009	0.0011	-0.0011	0.0628	0.0526	0.0476	0.0492
	15	-0.0049	-0.0010	0.0015	-0.0011	0.0628	0.0529	0.0477	0.0495
	20	-0.0049	-0.0008	0.0012	-0.0011	0.0622	0.0530	0.0480	0.0493
Off by 1.0	1	-0.0005	-0.0007	-0.0019	-0.0006	0.0687	0.0545	0.0515	0.0504
	3	-0.0014	-0.0021	-0.0028	0.0001	0.0618	0.0513	0.0491	0.0491
	5	-0.0005	-0.0016	-0.0015	0.0004	0.0622	0.0504	0.0480	0.0485
	10	-0.0008	-0.0011	-0.0022	0.0008	0.0608	0.0499	0.0479	0.0482
	15	-0.0007	-0.0014	-0.0019	0.0006	0.0603	0.0496	0.0478	0.0485
	20	-0.0009	-0.0015	-0.0021	0.0003	0.0600	0.0495	0.0481	0.0482

$\beta_3 = -1.00$ (Indigenous)									
On-target	1	0.0002	-0.0013	-0.0013	-0.0009	0.0351	0.0305	0.0263	0.0242
	3	0.0006	-0.0009	-0.0011	-0.0007	0.0330	0.0284	0.0254	0.0235
	5	0.0004	-0.0013	-0.0012	-0.0005	0.0324	0.0281	0.0253	0.0234
	10	0.0004	-0.0014	-0.0013	-0.0008	0.0322	0.0279	0.0251	0.0232
	15	0.0002	-0.0014	-0.0012	-0.0006	0.0319	0.0278	0.0249	0.0231
	20	0.0003	-0.0013	-0.0013	-0.0006	0.0319	0.0279	0.0248	0.0233
Off by 0.5	1	-0.0005	-0.0018	-0.0008	-0.0019	0.0355	0.0308	0.0261	0.0248
	3	-0.0005	-0.0012	-0.0009	-0.0017	0.0331	0.0293	0.0247	0.0240
	5	-0.0011	-0.0010	-0.0008	-0.0019	0.0331	0.0292	0.0249	0.0238
	10	-0.0009	-0.0013	-0.0008	-0.0018	0.0323	0.0289	0.0245	0.0237
	15	-0.0010	-0.0013	-0.0009	-0.0018	0.0324	0.0288	0.0245	0.0234
	20	-0.0007	-0.0012	-0.0010	-0.0017	0.0324	0.0288	0.0244	0.0234
Off by 1.0	1	-0.0007	-0.0019	-0.0009	0.0004	0.0341	0.0296	0.0269	0.0246
	3	-0.0007	-0.0020	-0.0015	0.0000	0.0321	0.0275	0.0258	0.0238
	5	-0.0006	-0.0015	-0.0014	-0.0002	0.0321	0.0273	0.0257	0.0237
	10	-0.0010	-0.0019	-0.0014	-0.0001	0.0317	0.0274	0.0256	0.0236
	15	-0.0005	-0.0020	-0.0015	-0.0002	0.0313	0.0271	0.0254	0.0235
	20	-0.0007	-0.0018	-0.0014	-0.0001	0.0311	0.0270	0.0254	0.0235

Table 26 Bias and RMSE of Parameter Estimates

Target	Number of PVs	Bias				RMSE			
		10	20	40	60	10	20	40	60
Mean									
On-target	1	0.0002	-0.0003	0.0006	0.0014	0.0346	0.0291	0.0254	0.0252
	3	0.0002	-0.0001	0.0005	0.0013	0.0328	0.0274	0.0248	0.0243
	5	0.0005	-0.0004	0.0005	0.0012	0.0318	0.0274	0.0244	0.0244
	10	0.0004	0.0000	0.0004	0.0011	0.0319	0.0272	0.0242	0.0242
	15	0.0001	-0.0002	0.0004	0.0010	0.0316	0.0269	0.0241	0.0241
	20	0.0003	-0.0001	0.0004	0.0011	0.0314	0.0269	0.0241	0.0241
Off by 0.5	1	-0.0002	-0.0012	0.0014	-0.0007	0.0348	0.0290	0.0272	0.0248
	3	-0.0003	-0.0007	0.0007	-0.0007	0.0328	0.0276	0.0257	0.0244
	5	-0.0005	-0.0005	0.0010	-0.0006	0.0325	0.0273	0.0254	0.0244
	10	-0.0005	-0.0005	0.0007	-0.0009	0.0318	0.0271	0.0253	0.0244
	15	-0.0003	-0.0005	0.0008	-0.0009	0.0319	0.0270	0.0252	0.0242
	20	-0.0004	-0.0006	0.0009	-0.0009	0.0318	0.0269	0.0252	0.0241
Off by 1.0	1	-0.0005	-0.0008	0.0001	-0.0014	0.0386	0.0318	0.0271	0.0261
	3	-0.0008	-0.0004	0.0006	-0.0013	0.0361	0.0301	0.0264	0.0260
	5	-0.0007	-0.0003	0.0008	-0.0010	0.0356	0.0295	0.0259	0.0255
	10	-0.0011	-0.0006	0.0008	-0.0011	0.0350	0.0293	0.0261	0.0253
	15	-0.0010	-0.0005	0.0007	-0.0013	0.0350	0.0293	0.0258	0.0253
	20	-0.0009	-0.0002	0.0007	-0.0013	0.0349	0.0292	0.0258	0.0253
Variance									
On-target	1	0.0831	0.0848	0.0839	0.0825	0.1517	0.1303	0.1174	0.1119
	3	0.0814	0.0842	0.0818	0.0828	0.1462	0.1248	0.1132	0.1106
	5	0.0808	0.0848	0.0817	0.0825	0.1449	0.1251	0.1133	0.1098
	10	0.0815	0.0854	0.0822	0.0825	0.1450	0.1252	0.1125	0.1093
	15	0.0824	0.0854	0.0817	0.0822	0.1450	0.1249	0.1122	0.1090
	20	0.0821	0.0851	0.0820	0.0821	0.1445	0.1245	0.1121	0.1088
Off by 0.5	1	0.0861	0.0833	0.0822	0.0854	0.1516	0.1300	0.1154	0.1163
	3	0.0849	0.0823	0.0818	0.0849	0.1469	0.1254	0.1128	0.1131
	5	0.0861	0.0808	0.0827	0.0849	0.1469	0.1243	0.1134	0.1125
	10	0.0850	0.0817	0.0821	0.0854	0.1451	0.1242	0.1128	0.1127
	15	0.0853	0.0817	0.0825	0.0850	0.1451	0.1239	0.1124	0.1120
	20	0.0845	0.0814	0.0834	0.0853	0.1447	0.1234	0.1132	0.1123
Off by 1.0	1	0.0848	0.0875	0.0818	0.0793	0.1529	0.1317	0.1176	0.1107
	3	0.0845	0.0876	0.0831	0.0801	0.1488	0.1280	0.1152	0.1099
	5	0.0832	0.0856	0.0821	0.0807	0.1477	0.1262	0.1141	0.1095
	10	0.0850	0.0867	0.0827	0.0804	0.1484	0.1265	0.1139	0.1093
	15	0.0829	0.0866	0.0826	0.0809	0.1461	0.1261	0.1137	0.1094
	20	0.0837	0.0868	0.0823	0.0804	0.1464	0.1264	0.1132	0.1091
Percentage in the Top Band									
On-target	1	0.0104	0.0104	0.0104	0.0104	0.0135	0.0130	0.0125	0.0124
	3	0.0100	0.0103	0.0103	0.0103	0.0127	0.0124	0.0119	0.0120
	5	0.0102	0.0102	0.0102	0.0103	0.0126	0.0121	0.0118	0.0119
	10	0.0102	0.0103	0.0102	0.0103	0.0125	0.0120	0.0117	0.0118
	15	0.0102	0.0103	0.0102	0.0103	0.0125	0.0121	0.0117	0.0118
	20	0.0102	0.0103	0.0102	0.0103	0.0124	0.0120	0.0117	0.0118
Off by 0.5	1	0.0103	0.0097	0.0106	0.0102	0.0132	0.0123	0.0126	0.0123
	3	0.0102	0.0100	0.0103	0.0103	0.0125	0.0119	0.0120	0.0120
	5	0.0103	0.0100	0.0105	0.0101	0.0124	0.0118	0.0121	0.0118
	10	0.0102	0.0101	0.0104	0.0101	0.0122	0.0118	0.0119	0.0117
	15	0.0102	0.0100	0.0104	0.0102	0.0121	0.0117	0.0119	0.0117
	20	0.0101	0.0101	0.0105	0.0101	0.0121	0.0117	0.0119	0.0117
Off by 1.0	1	0.0102	0.0104	0.0104	0.0095	0.0132	0.0128	0.0124	0.0117
	3	0.0102	0.0104	0.0104	0.0097	0.0122	0.0123	0.0120	0.0114
	5	0.0101	0.0103	0.0105	0.0098	0.0121	0.0120	0.0120	0.0115
	10	0.0102	0.0104	0.0105	0.0097	0.0121	0.0121	0.0119	0.0114
	15	0.0101	0.0104	0.0105	0.0097	0.0121	0.0120	0.0119	0.0113
	20	0.0101	0.0104	0.0104	0.0097	0.0120	0.0120	0.0119	0.0113

Table 27 Bias and RMSE of Estimated Parameters (Cont')

Target	Number of PVs	Bias				RMSE			
		10	20	40	60	10	20	40	60
90 th Percentile									
On-target	1	0.2101	0.2107	0.2106	0.2114	0.2242	0.2211	0.2183	0.2189
	3	0.2093	0.2110	0.2102	0.2125	0.2208	0.2187	0.2165	0.2183
	5	0.2094	0.2099	0.2098	0.2122	0.2198	0.2174	0.2156	0.2176
	10	0.2090	0.2106	0.2109	0.2116	0.2193	0.2176	0.2165	0.2169
	15	0.2090	0.2113	0.2104	0.2112	0.2191	0.2181	0.2159	0.2165
	20	0.2094	0.2113	0.2106	0.2112	0.2193	0.2181	0.2160	0.2165
Off by 0.5	1	0.2123	0.2095	0.2104	0.2101	0.2240	0.2194	0.2187	0.2186
	3	0.2103	0.2093	0.2087	0.2100	0.2199	0.2171	0.2152	0.2168
	5	0.2110	0.2088	0.2091	0.2103	0.2199	0.2163	0.2154	0.2169
	10	0.2104	0.2095	0.2096	0.2104	0.2188	0.2167	0.2155	0.2169
	15	0.2113	0.2092	0.2099	0.2099	0.2196	0.2163	0.2157	0.2161
	20	0.2110	0.2090	0.2098	0.2099	0.2192	0.2161	0.2155	0.2161
Off by 1.0	1	0.2085	0.2110	0.2106	0.2073	0.2204	0.2207	0.2188	0.2155
	3	0.2081	0.2120	0.2105	0.2076	0.2178	0.2194	0.2166	0.2141
	5	0.2094	0.2115	0.2101	0.2070	0.2181	0.2183	0.2161	0.2133
	10	0.2091	0.2113	0.2114	0.2065	0.2175	0.2179	0.2170	0.2128
	15	0.2085	0.2114	0.2106	0.2061	0.2167	0.2178	0.2162	0.2122
	20	0.2091	0.2110	0.2105	0.2062	0.2172	0.2175	0.2159	0.2123
97.5 th Percentile									
On-target	1	0.2521	0.2519	0.2555	0.2533	0.2778	0.2695	0.2719	0.2670
	3	0.2494	0.2505	0.2514	0.2533	0.2675	0.2633	0.2626	0.2639
	5	0.2507	0.2533	0.2535	0.2519	0.2681	0.2652	0.2638	0.2619
	10	0.2508	0.2539	0.2525	0.2527	0.2676	0.2652	0.2625	0.2620
	15	0.2513	0.2528	0.2533	0.2526	0.2674	0.2638	0.2630	0.2616
	20	0.2515	0.2524	0.2534	0.2519	0.2675	0.2633	0.2627	0.2608
Off by 0.5	1	0.2547	0.2555	0.2554	0.2557	0.2774	0.2736	0.2712	0.2703
	3	0.2509	0.2513	0.2529	0.2522	0.2677	0.2644	0.2641	0.2631
	5	0.2513	0.2513	0.2525	0.2537	0.2668	0.2639	0.2631	0.2645
	10	0.2518	0.2505	0.2515	0.2533	0.2662	0.2621	0.2616	0.2636
	15	0.2526	0.2520	0.2518	0.2533	0.2670	0.2636	0.2615	0.2633
	20	0.2523	0.2514	0.2530	0.2535	0.2663	0.2629	0.2626	0.2632
Off by 1.0	1	0.2536	0.2503	0.2493	0.2509	0.2763	0.2678	0.2651	0.2642
	3	0.2511	0.2532	0.2534	0.2489	0.2675	0.2662	0.2644	0.2607
	5	0.2504	0.2524	0.2510	0.2502	0.2652	0.2643	0.2620	0.2606
	10	0.2511	0.2525	0.2509	0.2495	0.2657	0.2635	0.2614	0.2594
	15	0.2492	0.2526	0.2515	0.2490	0.2635	0.2635	0.2618	0.2593
	20	0.2499	0.2531	0.2511	0.2494	0.2640	0.2636	0.2612	0.2594

Appendix 6 Standard Error Estimates of Simulation Two

Table 28 Bias and RMSE of the Generating Regression Coefficients

Target	On-target				Off by 0.5				Off by 1.0			
Number of PVs	10	20	40	60	10	20	40	60	10	20	40	60
$\beta_1 = -0.09$ (Gender)												
3	0.0050	0.0046	0.0044	0.0043	0.0050	0.0047	0.0044	0.0043	0.0050	0.0047	0.0044	0.0043
5	0.0049	0.0046	0.0043	0.0043	0.0049	0.0046	0.0044	0.0042	0.0050	0.0046	0.0044	0.0043
10	0.0049	0.0046	0.0043	0.0042	0.0048	0.0046	0.0043	0.0042	0.0049	0.0046	0.0043	0.0042
15	0.0048	0.0045	0.0043	0.0042	0.0048	0.0045	0.0043	0.0042	0.0049	0.0046	0.0043	0.0042
20	0.0048	0.0045	0.0043	0.0042	0.0048	0.0045	0.0043	0.0042	0.0048	0.0046	0.0043	0.0042
$\beta_2 = -0.57$ (LBOTE)												
3	0.0558	0.0516	0.0489	0.0477	0.0555	0.0519	0.0488	0.0476	0.0557	0.0519	0.0489	0.0478
5	0.0548	0.0512	0.0484	0.0473	0.0549	0.0511	0.0484	0.0473	0.0551	0.0513	0.0485	0.0472
10	0.0539	0.0506	0.0482	0.0470	0.0540	0.0506	0.0481	0.0470	0.0541	0.0507	0.0481	0.0470
15	0.0539	0.0506	0.0479	0.0470	0.0537	0.0505	0.0480	0.0469	0.0536	0.0504	0.0480	0.0470
20	0.0536	0.0504	0.0479	0.0469	0.0536	0.0503	0.0480	0.0469	0.0536	0.0507	0.0479	0.0469
$\beta_3 = -1.00$ (Indigenous)												
3	0.0280	0.0262	0.0245	0.0238	0.0280	0.0260	0.0243	0.0237	0.0276	0.0258	0.0243	0.0237
5	0.0275	0.0257	0.0242	0.0237	0.0274	0.0256	0.0241	0.0235	0.0271	0.0254	0.0242	0.0235
10	0.0270	0.0254	0.0240	0.0235	0.0270	0.0252	0.0240	0.0234	0.0269	0.0252	0.0240	0.0234
15	0.0269	0.0253	0.0239	0.0235	0.0267	0.0252	0.0239	0.0234	0.0268	0.0252	0.0239	0.0234
20	0.0268	0.0253	0.0240	0.0234	0.0267	0.0251	0.0239	0.0234	0.0268	0.0251	0.0238	0.0233

Table 29 Standard Error Estimates of the Population Parameter Estimates

Target Number of PVs	On-target				Off by 0.5				Off by 1.0			
	10	20	40	60	10	20	40	60	10	20	40	60
Mean												
3	0.0391	0.0378	0.0367	0.0363	0.0391	0.0377	0.0366	0.0362	0.0392	0.0378	0.0368	0.0363
5	0.0388	0.0375	0.0365	0.0362	0.0388	0.0375	0.0365	0.0362	0.0388	0.0377	0.0367	0.0362
10	0.0385	0.0373	0.0364	0.0361	0.0386	0.0373	0.0364	0.0361	0.0386	0.0374	0.0365	0.0361
15	0.0384	0.0372	0.0364	0.0360	0.0385	0.0373	0.0364	0.0360	0.0384	0.0374	0.0365	0.0361
20	0.0383	0.0372	0.0364	0.0360	0.0383	0.0372	0.0364	0.0360	0.0384	0.0373	0.0364	0.0360
Variance												
3	0.0950	0.0906	0.0861	0.0842	0.0949	0.0902	0.0860	0.0840	0.0954	0.0905	0.0862	0.0840
5	0.0944	0.0895	0.0856	0.0838	0.0939	0.0896	0.0853	0.0835	0.0934	0.0893	0.0854	0.0835
10	0.0923	0.0886	0.0849	0.0830	0.0926	0.0883	0.0848	0.0831	0.0925	0.0886	0.0851	0.0832
15	0.0922	0.0885	0.0846	0.0831	0.0921	0.0883	0.0846	0.0830	0.0925	0.0882	0.0848	0.0831
20	0.0920	0.0883	0.0846	0.0830	0.0918	0.0880	0.0845	0.0830	0.0923	0.0884	0.0846	0.0829
Percentage in the Top Band												
3	0.0109	0.0106	0.0103	0.0101	0.0110	0.0106	0.0102	0.0101	0.0109	0.0105	0.0102	0.0100
5	0.0108	0.0105	0.0101	0.0100	0.0107	0.0104	0.0101	0.0100	0.0108	0.0104	0.0101	0.0099
10	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099
15	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099
20	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099	0.0106	0.0103	0.0100	0.0099
90th Percentile												
3	0.0856	0.0838	0.0802	0.0789	0.0869	0.0835	0.0803	0.0787	0.0857	0.0829	0.0795	0.0782
5	0.0850	0.0823	0.0792	0.0780	0.0849	0.0816	0.0789	0.0776	0.0844	0.0816	0.0790	0.0774
10	0.0838	0.0813	0.0787	0.0776	0.0839	0.0808	0.0782	0.0771	0.0834	0.0805	0.0782	0.0768
15	0.0833	0.0811	0.0786	0.0773	0.0833	0.0806	0.0781	0.0769	0.0830	0.0805	0.0777	0.0764
20	0.0834	0.0809	0.0785	0.0771	0.0830	0.0804	0.0778	0.0768	0.0828	0.0805	0.0777	0.0764
97.5th Percentile												
3	0.1580	0.1531	0.1502	0.1479	0.1560	0.1532	0.1486	0.1474	0.1562	0.1516	0.1489	0.1457
5	0.1553	0.1526	0.1492	0.1465	0.1543	0.1515	0.1473	0.1458	0.1543	0.1500	0.1465	0.1448
10	0.1540	0.1512	0.1476	0.1456	0.1534	0.1501	0.1463	0.1448	0.1523	0.1491	0.1456	0.1440
15	0.1532	0.1503	0.1474	0.1457	0.1525	0.1496	0.1460	0.1445	0.1519	0.1487	0.1453	0.1436
20	0.1530	0.1500	0.1474	0.1453	0.1520	0.1494	0.1461	0.1444	0.1515	0.1480	0.1448	0.1435

Appendix 7 Results of Empirical Studies (NAPLAN 2015 Numeracy)

Table 30 Estimates of Mean from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	307549	0.1904	0.1901	0.1904	0.1904	0.1907	0.1906	0.1905	0.0000	-0.0004	-0.0001	0.0002	0.0001	0.0000
	NSW	97404	0.2360	0.2360	0.2345	0.2349	0.2348	0.2353	0.2349	0.0015	0.0016	0.0004	0.0003	0.0009	0.0005
	VIC	74689	0.3438	0.3437	0.3449	0.3448	0.3451	0.3453	0.3455	-0.0012	-0.0012	-0.0001	0.0002	0.0003	0.0006
	QLD	65420	0.1156	0.1157	0.1170	0.1167	0.1173	0.1163	0.1162	-0.0013	-0.0013	-0.0003	0.0003	-0.0007	-0.0007
	WA	34152	0.0753	0.0699	0.0734	0.0745	0.0738	0.0731	0.0728	0.0019	-0.0035	0.0011	0.0004	-0.0003	-0.0006
	SA	20570	-0.0145	-0.0127	-0.0133	-0.0140	-0.0128	-0.0121	-0.0123	-0.0012	0.0007	-0.0006	0.0005	0.0012	0.0010
	TAS	6584	0.1061	0.1031	0.1061	0.1011	0.1041	0.1019	0.1041	0.0000	-0.0030	-0.0050	-0.0020	-0.0042	-0.0020
	ACT	5229	0.3530	0.3534	0.3482	0.3456	0.3480	0.3476	0.3481	0.0048	0.0052	-0.0026	-0.0002	-0.0006	0.0000
5	NT	3501	-0.6698	-0.6572	-0.6582	-0.6630	-0.6570	-0.6610	-0.6619	-0.0116	0.0010	-0.0048	0.0012	-0.0027	-0.0037
	Aus	286154	0.3074	0.3067	0.3064	0.3067	0.3068	0.3067	0.3066	0.0010	0.0003	0.0003	0.0004	0.0003	0.0002
	NSW	90745	0.3879	0.3853	0.3861	0.3872	0.3869	0.3865	0.3869	0.0017	-0.0008	0.0010	0.0008	0.0004	0.0008
	VIC	69171	0.4864	0.4887	0.4866	0.4866	0.4862	0.4866	0.4864	-0.0002	0.0021	0.0000	-0.0003	0.0000	-0.0002
	QLD	61835	0.1951	0.1955	0.1952	0.1957	0.1963	0.1956	0.1954	-0.0001	0.0003	0.0005	0.0010	0.0004	0.0002
	WA	31266	0.1744	0.1749	0.1737	0.1733	0.1746	0.1748	0.1744	0.0006	0.0012	-0.0004	0.0009	0.0011	0.0007
	SA	19088	0.0547	0.0522	0.0543	0.0533	0.0539	0.0537	0.0538	0.0004	-0.0022	-0.0011	-0.0004	-0.0006	-0.0006
	TAS	6008	0.1461	0.1432	0.1426	0.1436	0.1400	0.1432	0.1406	0.0035	0.0006	0.0010	-0.0026	0.0006	-0.0020
7	ACT	4786	0.4735	0.4632	0.4662	0.4656	0.4688	0.4660	0.4663	0.0073	-0.0031	-0.0006	0.0026	-0.0002	0.0001
	NT	3255	-0.7492	-0.7632	-0.7685	-0.7662	-0.7693	-0.7694	-0.7675	0.0193	0.0053	0.0024	-0.0008	-0.0008	0.0010
	Aus	278343	-0.1056	-0.1050	-0.1053	-0.1050	-0.1055	-0.1054	-0.1053	-0.0002	0.0003	0.0003	-0.0002	-0.0001	0.0000
	NSW	88648	-0.0352	-0.0354	-0.0340	-0.0351	-0.0354	-0.0350	-0.0351	-0.0012	-0.0014	-0.0011	-0.0015	-0.0010	-0.0011
	VIC	68159	-0.0140	-0.0129	-0.0140	-0.0126	-0.0138	-0.0129	-0.0133	0.0000	0.0010	0.0013	0.0002	0.0011	0.0007
	QLD	57996	-0.1643	-0.1631	-0.1656	-0.1641	-0.1645	-0.1653	-0.1649	0.0013	0.0025	0.0015	0.0011	0.0003	0.0007
	WA	30065	-0.1735	-0.1730	-0.1740	-0.1735	-0.1739	-0.1738	-0.1733	0.0005	0.0010	0.0004	0.0000	0.0001	0.0006
	SA	18974	-0.2670	-0.2658	-0.2638	-0.2649	-0.2650	-0.2652	-0.2645	-0.0032	-0.0020	-0.0012	-0.0012	-0.0014	-0.0007
9	TAS	6292	-0.3318	-0.3287	-0.3303	-0.3300	-0.3270	-0.3309	-0.3293	-0.0016	0.0016	0.0003	0.0033	-0.0007	0.0010
	ACT	4886	0.0045	0.0029	0.0040	0.0058	0.0052	0.0045	0.0055	0.0005	-0.0011	0.0018	0.0011	0.0004	0.0014
	NT	3323	-1.0332	-1.0420	-1.0404	-1.0404	-1.0413	-1.0412	-1.0400	0.0072	-0.0017	0.0000	-0.0010	-0.0008	0.0004
	Aus	281602	-0.3216	-0.3221	-0.3218	-0.3225	-0.3224	-0.3225	-0.3224	0.0002	-0.0003	-0.0007	-0.0006	-0.0007	-0.0006
	NSW	89624	-0.2481	-0.2491	-0.2469	-0.2489	-0.2484	-0.2479	-0.2484	-0.0012	-0.0023	-0.0020	-0.0016	-0.0011	-0.0015
	VIC	68446	-0.2350	-0.2334	-0.2361	-0.2362	-0.2360	-0.2363	-0.2357	0.0011	0.0028	-0.0001	0.0001	-0.0002	0.0004
	QLD	59293	-0.4388	-0.4408	-0.4392	-0.4390	-0.4389	-0.4398	-0.4395	0.0005	-0.0015	0.0003	0.0003	-0.0005	-0.0002
	WA	30256	-0.2520	-0.2507	-0.2525	-0.2504	-0.2514	-0.2513	-0.2510	0.0005	0.0019	0.0021	0.0011	0.0012	0.0015
9	SA	19451	-0.5363	-0.5371	-0.5367	-0.5395	-0.5380	-0.5386	-0.5388	0.0005	-0.0004	-0.0028	-0.0013	-0.0019	-0.0021
	TAS	6512	-0.5540	-0.5627	-0.5622	-0.5656	-0.5677	-0.5649	-0.5663	0.0082	-0.0005	-0.0034	-0.0055	-0.0027	-0.0041
	ACT	4997	-0.1922	-0.1903	-0.1822	-0.1845	-0.1883	-0.1908	-0.1889	-0.0100	-0.0081	-0.0024	-0.0061	-0.0086	-0.0068
	NT	3023	-1.1833	-1.1907	-1.1962	-1.1965	-1.1953	-1.1938	-1.1945	0.0129	0.0055	-0.0002	0.0009	0.0024	0.0017

Table 31 Estimates of Standard Deviation from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	307549	0.9748	0.9749	0.9745	0.9747	0.9750	0.9750	0.9748	0.0003	0.0004	0.0002	0.0005	0.0005	0.0003
	NSW	97404	0.9997	0.9999	0.9996	1.0007	1.0004	1.0004	1.0000	0.0001	0.0003	0.0012	0.0008	0.0008	0.0005
	VIC	74689	0.9476	0.9492	0.9489	0.9483	0.9491	0.9488	0.9488	-0.0013	0.0003	-0.0006	0.0002	-0.0001	-0.0001
	QLD	65420	0.9252	0.9270	0.9267	0.9268	0.9273	0.9276	0.9267	-0.0015	0.0002	0.0000	0.0005	0.0008	-0.0001
	WA	34152	0.9767	0.9764	0.9763	0.9756	0.9768	0.9769	0.9766	0.0004	0.0001	-0.0007	0.0005	0.0006	0.0004
	SA	20570	0.9257	0.9209	0.9182	0.9179	0.9202	0.9195	0.9195	0.0075	0.0027	-0.0003	0.0020	0.0013	0.0013
	TAS	6584	0.9664	0.9619	0.9620	0.9592	0.9580	0.9577	0.9607	0.0043	-0.0002	-0.0028	-0.0040	-0.0044	-0.0014
	ACT	5229	0.9315	0.9290	0.9292	0.9316	0.9298	0.9300	0.9295	0.0023	-0.0002	0.0024	0.0006	0.0008	0.0003
5	NT	3501	1.1344	1.1271	1.1286	1.1302	1.1232	1.1282	1.1282	0.0058	-0.0015	0.0016	-0.0054	-0.0005	-0.0004
	Aus	286154	1.1753	1.1763	1.1766	1.1764	1.1763	1.1763	1.1766	-0.0013	-0.0003	-0.0002	-0.0003	-0.0003	-0.0001
	NSW	90745	1.2383	1.2405	1.2391	1.2402	1.2405	1.2398	1.2402	-0.0008	0.0014	0.0011	0.0014	0.0008	0.0012
	VIC	69171	1.1260	1.1242	1.1261	1.1256	1.1254	1.1262	1.1259	-0.0001	-0.0019	-0.0006	-0.0007	0.0001	-0.0002
	QLD	61835	1.0965	1.0971	1.0960	1.0957	1.0951	1.0948	1.0958	0.0005	0.0011	-0.0003	-0.0009	-0.0012	-0.0002
	WA	31266	1.1756	1.1741	1.1774	1.1745	1.1750	1.1748	1.1748	-0.0018	-0.0033	-0.0029	-0.0024	-0.0026	-0.0026
	SA	19088	1.0712	1.0722	1.0763	1.0759	1.0747	1.0755	1.0752	-0.0051	-0.0041	-0.0004	-0.0016	-0.0009	-0.0012
	TAS	6008	1.1228	1.1304	1.1297	1.1266	1.1290	1.1291	1.1288	-0.0069	0.0006	-0.0032	-0.0007	-0.0007	-0.0009
7	ACT	4786	1.0881	1.0981	1.0953	1.0935	1.0940	1.0932	1.0960	-0.0073	0.0028	-0.0018	-0.0013	-0.0021	0.0007
	NT	3255	1.4005	1.4149	1.4179	1.4209	1.4173	1.4193	1.4178	-0.0174	-0.0031	0.0030	-0.0006	0.0013	-0.0002
	Aus	278343	1.1150	1.1158	1.1154	1.1154	1.1153	1.1157	1.1154	-0.0004	0.0004	0.0000	-0.0001	0.0003	0.0000
	NSW	88648	1.2119	1.2126	1.2126	1.2127	1.2133	1.2130	1.2129	-0.0008	-0.0001	0.0001	0.0006	0.0003	0.0003
	VIC	68159	1.0732	1.0740	1.0738	1.0721	1.0728	1.0732	1.0726	-0.0006	0.0002	-0.0016	-0.0009	-0.0006	-0.0012
	QLD	57996	1.0224	1.0240	1.0226	1.0236	1.0227	1.0236	1.0228	-0.0002	0.0014	0.0010	0.0001	0.0011	0.0003
	WA	30065	1.0909	1.0910	1.0910	1.0915	1.0918	1.0914	1.0920	0.0000	0.0000	0.0005	0.0008	0.0004	0.0010
	SA	18974	0.9804	0.9808	0.9802	0.9813	0.9782	0.9815	0.9806	0.0002	0.0006	0.0011	-0.0020	0.0013	0.0003
9	TAS	6292	0.9812	0.9818	0.9809	0.9805	0.9802	0.9811	0.9805	0.0003	0.0009	-0.0004	-0.0008	0.0001	-0.0004
	ACT	4886	1.0826	1.0772	1.0762	1.0769	1.0746	1.0741	1.0750	0.0064	0.0010	0.0007	-0.0017	-0.0022	-0.0013
	NT	3323	1.3186	1.3276	1.3209	1.3231	1.3212	1.3211	1.3243	-0.0023	0.0067	0.0022	0.0003	0.0002	0.0034
	Aus	281602	1.1464	1.1470	1.1471	1.1473	1.1471	1.1473	1.1471	-0.0007	-0.0001	0.0002	0.0000	0.0002	0.0000
	NSW	89624	1.2486	1.2493	1.2480	1.2494	1.2491	1.2492	1.2488	0.0006	0.0012	0.0013	0.0010	0.0012	0.0007
	VIC	68446	1.1311	1.1320	1.1327	1.1324	1.1323	1.1322	1.1325	-0.0016	-0.0007	-0.0002	-0.0004	-0.0004	-0.0002
	QLD	59293	1.0174	1.0172	1.0172	1.0169	1.0173	1.0173	1.0168	0.0002	0.0000	-0.0004	0.0001	0.0001	-0.0004
	WA	30256	1.1191	1.1209	1.1228	1.1225	1.1220	1.1228	1.1221	-0.0038	-0.0020	-0.0003	-0.0008	0.0000	-0.0008
9	SA	19451	1.0283	1.0280	1.0282	1.0276	1.0264	1.0278	1.0288	0.0001	-0.0002	-0.0006	-0.0018	-0.0004	0.0006
	TAS	6512	1.0198	1.0166	1.0181	1.0169	1.0166	1.0173	1.0161	0.0017	-0.0015	-0.0012	-0.0015	-0.0008	-0.0020
	ACT	4997	1.1246	1.1225	1.1251	1.1227	1.1256	1.1220	1.1237	-0.0005	-0.0026	-0.0024	0.0005	-0.0031	-0.0014
	NT	3023	1.1945	1.1913	1.1958	1.1928	1.1945	1.1958	1.1954	-0.0013	-0.0045	-0.0030	-0.0013	-0.0001	-0.0004

Table 32 Estimates of Percentage in the Top Band from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	307549	19.7160	19.7116	19.6904	19.7100	19.7039	19.7166	19.6963	0.0256	0.0212	0.0196	0.0135	0.0263	0.0059
	NSW	97404	21.7984	21.7476	21.6770	21.7415	21.6947	21.7326	21.6883	0.1214	0.0707	0.0645	0.0177	0.0556	0.0113
	VIC	74689	23.5157	23.5455	23.5278	23.5628	23.5682	23.6178	23.5862	-0.0121	0.0177	0.0351	0.0405	0.0901	0.0584
	QLD	65420	16.5479	16.6563	16.6241	16.5978	16.6438	16.6147	16.6140	-0.0762	0.0322	-0.0263	0.0197	-0.0094	-0.0101
	WA	34152	16.6588	16.6134	16.6937	16.7583	16.6833	16.7001	16.6572	-0.0349	-0.0804	0.0646	-0.0105	0.0064	-0.0366
	SA	20570	13.3976	13.2055	13.1939	13.1531	13.2810	13.2406	13.2663	0.2037	0.0116	-0.0407	0.0871	0.0467	0.0725
	TAS	6584	17.8207	17.1922	17.4652	17.3261	17.3498	17.1082	17.3496	0.3555	-0.2731	-0.1391	-0.1154	-0.3570	-0.1156
	ACT	5229	23.2839	23.9600	23.7715	23.4068	23.4399	23.3385	23.3590	-0.4875	0.1885	-0.3647	-0.3315	-0.4329	-0.4125
5	NT	3501	6.0756	6.1822	6.3140	6.4012	6.0853	6.3241	6.2418	-0.2384	-0.1318	0.0872	-0.2287	0.0102	-0.0722
	Aus	286154	19.6128	19.6308	19.6498	19.6599	19.6481	19.6313	19.6406	-0.0369	-0.0189	0.0101	-0.0016	-0.0185	-0.0092
	NSW	90745	22.4255	22.5001	22.5054	22.5202	22.5186	22.5039	22.5182	-0.0799	-0.0053	0.0148	0.0132	-0.0015	0.0128
	VIC	69171	23.1494	23.1320	23.1488	23.1732	23.0856	23.1099	23.1175	0.0006	-0.0168	0.0244	-0.0632	-0.0388	-0.0312
	QLD	61835	15.8892	15.8995	15.8643	15.9867	15.9590	15.8929	15.9240	0.0249	0.0353	0.1224	0.0947	0.0286	0.0598
	WA	31266	16.9527	16.9030	16.9624	16.8704	16.9643	16.9525	16.9223	-0.0097	-0.0594	-0.0920	0.0019	-0.0099	-0.0401
	SA	19088	12.6687	12.6419	12.8164	12.6933	12.7061	12.7672	12.7132	-0.1478	-0.1746	-0.1232	-0.1103	-0.0493	-0.1032
	TAS	6008	15.8757	15.9939	16.1696	15.8521	15.9399	15.8943	15.8375	-0.2939	-0.1756	-0.3175	-0.2297	-0.2753	-0.3321
7	ACT	4786	22.1346	22.3406	22.2539	22.2092	22.3477	22.1400	22.2810	-0.1193	0.0866	-0.0447	0.0937	-0.1140	0.0270
	NT	3255	7.2078	6.9884	7.0323	6.9226	6.9550	6.7816	6.9503	0.1755	-0.0439	-0.1097	-0.0773	-0.2507	-0.0820
	Aus	278343	18.9583	19.0115	19.0039	19.0124	19.0056	19.0110	18.9998	-0.0455	0.0076	0.0086	0.0018	0.0071	-0.0041
	NSW	88648	21.5849	21.6356	21.6405	21.6391	21.6472	21.6631	21.6392	-0.0556	-0.0049	-0.0014	0.0067	0.0226	-0.0013
	VIC	68159	20.7751	20.7536	20.6869	20.7091	20.7351	20.7321	20.7109	0.0882	0.0667	0.0222	0.0482	0.0452	0.0241
	QLD	57996	16.4817	16.6079	16.6034	16.6449	16.5838	16.5712	16.5760	-0.1217	0.0046	0.0416	-0.0195	-0.0322	-0.0273
	WA	30065	17.4957	17.6574	17.6345	17.5876	17.5936	17.6239	17.6214	-0.1388	0.0229	-0.0468	-0.0409	-0.0106	-0.0130
	SA	18974	13.3774	13.4508	13.5546	13.5111	13.4944	13.4971	13.4816	-0.1772	-0.1038	-0.0435	-0.0601	-0.0575	-0.0730
9	TAS	6292	12.3851	12.5249	12.5915	12.5528	12.5797	12.5746	12.6102	-0.2064	-0.0667	-0.0387	-0.0118	-0.0169	0.0187
	ACT	4886	22.0625	21.7292	21.7500	21.9000	21.7486	21.7813	21.7668	0.3125	-0.0208	0.1500	-0.0014	0.0313	0.0168
	NT	3323	7.8014	7.4622	7.5979	7.6997	7.6616	7.6364	7.7552	0.2035	-0.1357	0.1018	0.0637	0.0385	0.1573
	Aus	281602	18.7912	18.6934	18.7097	18.7005	18.6766	18.7017	18.6977	0.0816	-0.0163	-0.0091	-0.0331	-0.0080	-0.0120
	NSW	89624	21.4441	21.3731	21.4792	21.4127	21.3845	21.4090	21.3996	-0.0351	-0.1061	-0.0665	-0.0947	-0.0702	-0.0796
	VIC	68446	20.6452	20.5304	20.5125	20.4631	20.4511	20.5089	20.5078	0.1327	0.0179	-0.0494	-0.0614	-0.0037	-0.0048
	QLD	59293	14.6092	14.3985	14.4191	14.4845	14.4732	14.4690	14.4483	0.1901	-0.0206	0.0654	0.0541	0.0499	0.0293
	WA	30256	20.7629	20.7228	20.6177	20.7428	20.7219	20.7183	20.7335	0.1452	0.1051	0.1251	0.1042	0.1006	0.1158
9	SA	19451	12.8384	12.7894	12.7765	12.6464	12.6576	12.7303	12.7369	0.0619	0.0129	-0.1301	-0.1189	-0.0462	-0.0396
	TAS	6512	12.6905	12.7475	12.6252	12.6656	12.4852	12.5925	12.5996	0.0653	0.1223	0.0404	-0.1400	-0.0327	-0.0256
	ACT	4997	22.1561	22.0397	22.0821	22.3306	22.1328	21.9938	22.0262	0.0739	-0.0424	0.2485	0.0507	-0.0883	-0.0560
	NT	3023	6.4931	7.0003	6.6960	6.7264	6.6824	6.6368	6.7092	-0.2029	0.3044	0.0304	-0.0135	-0.0592	0.0132

Table 33 Estimates of 20th Percentile from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	307549	-0.6237	-0.6252	-0.6253	-0.6242	-0.6243	-0.6241	-0.6240	0.0016	0.0001	0.0011	0.0010	0.0012	0.0013
	NSW	97404	-0.6094	-0.6112	-0.6136	-0.6121	-0.6106	-0.6116	-0.6108	0.0042	0.0024	0.0015	0.0031	0.0021	0.0029
	VIC	74689	-0.4574	-0.4561	-0.4563	-0.4546	-0.4550	-0.4553	-0.4548	-0.0011	0.0001	0.0017	0.0012	0.0009	0.0015
	QLD	65420	-0.6612	-0.6626	-0.6618	-0.6613	-0.6609	-0.6625	-0.6620	0.0006	-0.0008	0.0005	0.0009	-0.0007	-0.0002
	WA	34152	-0.7295	-0.7430	-0.7382	-0.7345	-0.7390	-0.7375	-0.7382	0.0087	-0.0048	0.0037	-0.0007	0.0008	0.0001
	SA	20570	-0.7952	-0.7811	-0.7824	-0.7827	-0.7846	-0.7802	-0.7812	-0.0128	0.0013	-0.0003	-0.0022	0.0023	0.0012
	TAS	6584	-0.7108	-0.7121	-0.7055	-0.7138	-0.7092	-0.7088	-0.7055	-0.0053	-0.0066	-0.0083	-0.0037	-0.0033	0.0000
	ACT	5229	-0.4201	-0.4270	-0.4425	-0.4324	-0.4263	-0.4297	-0.4338	0.0225	0.0155	0.0101	0.0162	0.0128	0.0087
5	NT	3501	-1.6374	-1.5895	-1.6133	-1.6151	-1.6056	-1.6114	-1.6116	-0.0241	0.0238	-0.0018	0.0077	0.0019	0.0018
	Aus	286154	-0.6807	-0.6826	-0.6826	-0.6825	-0.6817	-0.6824	-0.6827	0.0020	0.0000	0.0002	0.0009	0.0002	-0.0001
	NSW	90745	-0.6648	-0.6705	-0.6688	-0.6684	-0.6690	-0.6699	-0.6698	0.0040	-0.0016	0.0005	-0.0001	-0.0010	-0.0010
	VIC	69171	-0.4777	-0.4723	-0.4750	-0.4744	-0.4751	-0.4754	-0.4761	-0.0027	0.0028	0.0006	-0.0001	-0.0004	-0.0011
	QLD	61835	-0.7355	-0.7328	-0.7311	-0.7311	-0.7296	-0.7301	-0.7304	-0.0043	-0.0016	0.0000	0.0015	0.0010	0.0008
	WA	31266	-0.8074	-0.8113	-0.8117	-0.8135	-0.8116	-0.8119	-0.8128	0.0043	0.0004	-0.0018	0.0001	-0.0001	-0.0011
	SA	19088	-0.8467	-0.8434	-0.8485	-0.8511	-0.8494	-0.8495	-0.8481	0.0018	0.0051	-0.0026	-0.0009	-0.0010	0.0004
	TAS	6008	-0.8092	-0.8346	-0.8246	-0.8232	-0.8275	-0.8230	-0.8249	0.0154	-0.0100	0.0014	-0.0029	0.0017	-0.0003
7	ACT	4786	-0.4485	-0.4952	-0.4728	-0.4804	-0.4776	-0.4751	-0.4815	0.0243	-0.0225	-0.0077	-0.0049	-0.0024	-0.0087
	NT	3255	-1.9838	-1.9996	-1.9940	-1.9975	-1.9987	-2.0062	-1.9957	0.0102	-0.0055	-0.0034	-0.0047	-0.0122	-0.0017
	Aus	278343	-1.0422	-1.0420	-1.0426	-1.0424	-1.0423	-1.0426	-1.0416	0.0004	0.0006	0.0002	0.0003	0.0000	0.0010
	NSW	88648	-1.0594	-1.0565	-1.0551	-1.0570	-1.0584	-1.0569	-1.0565	-0.0043	-0.0014	-0.0019	-0.0033	-0.0019	-0.0015
	VIC	68159	-0.9263	-0.9271	-0.9285	-0.9271	-0.9272	-0.9265	-0.9263	0.0022	0.0014	0.0014	0.0013	0.0020	0.0022
	QLD	57996	-1.0390	-1.0402	-1.0427	-1.0404	-1.0412	-1.0404	-1.0394	0.0037	0.0025	0.0023	0.0015	0.0023	0.0033
	WA	30065	-1.1044	-1.0987	-1.1071	-1.1060	-1.1042	-1.1054	-1.1038	0.0028	0.0084	0.0011	0.0029	0.0017	0.0033
	SA	18974	-1.0934	-1.0926	-1.0915	-1.0946	-1.0907	-1.0943	-1.0909	-0.0020	-0.0011	-0.0031	0.0008	-0.0028	0.0006
9	TAS	6292	-1.1780	-1.1785	-1.1746	-1.1767	-1.1701	-1.1759	-1.1719	-0.0035	-0.0039	-0.0021	0.0045	-0.0013	0.0027
	ACT	4886	-0.9262	-0.9279	-0.9187	-0.9195	-0.9192	-0.9218	-0.9182	-0.0075	-0.0092	-0.0008	-0.0005	-0.0031	0.0005
	NT	3323	-2.0346	-2.0339	-2.0303	-2.0372	-2.0397	-2.0376	-2.0382	-0.0043	-0.0036	-0.0069	-0.0093	-0.0073	-0.0079
	Aus	281602	-1.2807	-1.2809	-1.2812	-1.2820	-1.2821	-1.2817	-1.2818	0.0005	0.0002	-0.0008	-0.0010	-0.0005	-0.0006
	NSW	89624	-1.2893	-1.2905	-1.2904	-1.2907	-1.2901	-1.2897	-1.2897	0.0011	-0.0001	-0.0003	0.0003	0.0007	0.0007
	VIC	68446	-1.1859	-1.1890	-1.1911	-1.1908	-1.1907	-1.1917	-1.1913	0.0051	0.0021	0.0002	0.0003	-0.0007	-0.0003
	QLD	59293	-1.3055	-1.3038	-1.3020	-1.3039	-1.3035	-1.3030	-1.3032	-0.0035	-0.0018	-0.0018	-0.0015	-0.0010	-0.0012
	WA	30256	-1.2082	-1.2113	-1.2131	-1.2113	-1.2138	-1.2118	-1.2112	0.0049	0.0018	0.0018	-0.0007	0.0013	0.0020
9	SA	19451	-1.3963	-1.3944	-1.3954	-1.3976	-1.3978	-1.4000	-1.3994	-0.0009	0.0010	-0.0022	-0.0024	-0.0046	-0.0040
	TAS	6512	-1.4342	-1.4389	-1.4388	-1.4390	-1.4416	-1.4405	-1.4370	0.0046	-0.0001	-0.0002	-0.0027	-0.0016	0.0019
	ACT	4997	-1.1577	-1.1397	-1.1420	-1.1374	-1.1483	-1.1444	-1.1450	-0.0157	0.0024	0.0046	-0.0063	-0.0024	-0.0030
	NT	3023	-2.1498	-2.1421	-2.1458	-2.1559	-2.1555	-2.1494	-2.1532	-0.0040	0.0037	-0.0101	-0.0097	-0.0035	-0.0074

Table 34 Estimates of 80th Percentile from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	307549	0.9966	0.9962	0.9953	0.9959	0.9959	0.9963	0.9956	0.0013	0.0009	0.0006	0.0006	0.0010	0.0003
	NSW	97404	1.0725	1.0703	1.0692	1.0713	1.0699	1.0706	1.0692	0.0033	0.0011	0.0021	0.0007	0.0014	0.0000
	VIC	74689	1.1279	1.1264	1.1272	1.1286	1.1285	1.1292	1.1302	0.0006	-0.0008	0.0014	0.0013	0.0020	0.0030
	QLD	65420	0.8838	0.8819	0.8826	0.8814	0.8833	0.8826	0.8826	0.0012	-0.0007	-0.0012	0.0007	0.0000	0.0000
	WA	34152	0.8788	0.8774	0.8827	0.8835	0.8804	0.8805	0.8790	-0.0039	-0.0053	0.0008	-0.0023	-0.0022	-0.0037
	SA	20570	0.7537	0.7413	0.7454	0.7430	0.7471	0.7460	0.7454	0.0083	-0.0042	-0.0024	0.0016	0.0006	-0.0001
	TAS	6584	0.9250	0.9076	0.9097	0.9048	0.9096	0.8991	0.9072	0.0153	-0.0021	-0.0048	-0.0001	-0.0106	-0.0025
	ACT	5229	1.1209	1.1314	1.1268	1.1211	1.1222	1.1168	1.1182	-0.0058	0.0047	-0.0057	-0.0046	-0.0099	-0.0086
5	NT	3501	0.2949	0.2833	0.2889	0.2811	0.2795	0.2825	0.2795	0.0060	-0.0057	-0.0078	-0.0094	-0.0065	-0.0095
	Aus	286154	1.2748	1.2759	1.2768	1.2769	1.2769	1.2762	1.2765	-0.0020	-0.0009	0.0001	0.0001	-0.0006	-0.0003
	NSW	90745	1.4078	1.4114	1.4100	1.4135	1.4130	1.4121	1.4122	-0.0022	0.0013	0.0035	0.0030	0.0021	0.0022
	VIC	69171	1.4233	1.4231	1.4238	1.4246	1.4220	1.4213	1.4228	-0.0005	-0.0007	0.0008	-0.0018	-0.0025	-0.0010
	QLD	61835	1.1125	1.1154	1.1102	1.1136	1.1132	1.1113	1.1123	0.0023	0.0052	0.0033	0.0029	0.0010	0.0021
	WA	31266	1.1498	1.1455	1.1506	1.1476	1.1499	1.1524	1.1504	-0.0008	-0.0051	-0.0030	-0.0007	0.0018	-0.0002
	SA	19088	0.9378	0.9413	0.9527	0.9474	0.9440	0.9483	0.9447	-0.0149	-0.0114	-0.0053	-0.0087	-0.0044	-0.0080
	TAS	6008	1.1126	1.1172	1.1143	1.1041	1.1086	1.1056	1.1019	-0.0016	0.0029	-0.0101	-0.0057	-0.0087	-0.0123
7	ACT	4786	1.3908	1.3902	1.3851	1.3812	1.3901	1.3776	1.3844	0.0057	0.0050	-0.0039	0.0050	-0.0075	-0.0007
	NT	3255	0.4502	0.4478	0.4475	0.4505	0.4542	0.4538	0.4498	0.0027	0.0003	0.0030	0.0068	0.0063	0.0023
	Aus	278343	0.7795	0.7813	0.7798	0.7808	0.7808	0.7805	0.7803	-0.0003	0.0015	0.0010	0.0010	0.0007	0.0005
	NSW	88648	0.9100	0.9124	0.9109	0.9123	0.9111	0.9121	0.9115	-0.0008	0.0015	0.0014	0.0002	0.0012	0.0006
	VIC	68159	0.8615	0.8617	0.8582	0.8595	0.8607	0.8606	0.8598	0.0033	0.0035	0.0013	0.0025	0.0024	0.0016
	QLD	57996	0.6702	0.6794	0.6758	0.6817	0.6786	0.6775	0.6774	-0.0056	0.0036	0.0059	0.0028	0.0016	0.0015
	WA	30065	0.7174	0.7168	0.7175	0.7131	0.7185	0.7178	0.7183	-0.0001	-0.0007	-0.0044	0.0010	0.0003	0.0008
	SA	18974	0.5253	0.5187	0.5234	0.5197	0.5217	0.5208	0.5202	0.0020	-0.0047	-0.0036	-0.0016	-0.0026	-0.0032
9	TAS	6292	0.4668	0.4701	0.4778	0.4821	0.4803	0.4797	0.4763	-0.0110	-0.0077	0.0043	0.0025	0.0019	-0.0014
	ACT	4886	0.9228	0.9111	0.9067	0.9102	0.9042	0.9095	0.9077	0.0162	0.0044	0.0035	-0.0025	0.0029	0.0010
	NT	3323	0.0252	0.0486	0.0284	0.0417	0.0265	0.0317	0.0350	-0.0033	0.0202	0.0133	-0.0019	0.0033	0.0065
	Aus	281602	0.5795	0.5770	0.5773	0.5768	0.5760	0.5770	0.5767	0.0022	-0.0003	-0.0005	-0.0013	-0.0003	-0.0006
	NSW	89624	0.7141	0.7125	0.7148	0.7122	0.7116	0.7130	0.7118	-0.0006	-0.0023	-0.0026	-0.0031	-0.0018	-0.0029
	VIC	68446	0.6690	0.6642	0.6617	0.6616	0.6607	0.6636	0.6630	0.0072	0.0025	-0.0001	-0.0010	0.0019	0.0012
	QLD	59293	0.3861	0.3849	0.3814	0.3876	0.3864	0.3857	0.3862	0.0047	0.0035	0.0063	0.0050	0.0043	0.0048
	WA	30256	0.6751	0.6734	0.6687	0.6719	0.6718	0.6717	0.6724	0.0063	0.0047	0.0031	0.0030	0.0030	0.0036
9	SA	19451	0.2781	0.2886	0.2878	0.2840	0.2868	0.2868	0.2866	-0.0098	0.0008	-0.0039	-0.0010	-0.0011	-0.0012
	TAS	6512	0.3001	0.2846	0.2832	0.2808	0.2752	0.2782	0.2755	0.0169	0.0014	-0.0023	-0.0080	-0.0050	-0.0077
	ACT	4997	0.7351	0.7197	0.7308	0.7379	0.7284	0.7237	0.7255	0.0044	-0.0111	0.0071	-0.0024	-0.0071	-0.0053
	NT	3023	-0.2292	-0.2289	-0.2374	-0.2372	-0.2340	-0.2271	-0.2300	0.0082	0.0085	0.0002	0.0034	0.0103	0.0074

Table 35 Standard Error Estimates of Mean from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aus	307549	0.0069	0.0068	0.0068	0.0068	0.0068	0.0068	0.0001	0.0000	0.0001	0.0000	0.0000
	NSW	97404	0.0128	0.0128	0.0128	0.0128	0.0128	0.0128	0.0000	0.0000	0.0000	0.0000	0.0000
	VIC	74689	0.0122	0.0122	0.0122	0.0122	0.0123	0.0122	0.0000	0.0000	0.0000	0.0001	0.0001
	QLD	65420	0.0139	0.0139	0.0139	0.0139	0.0140	0.0139	0.0001	0.0000	0.0001	0.0002	0.0000
	WA	34152	0.0191	0.0190	0.0190	0.0191	0.0191	0.0190	0.0001	0.0000	0.0002	0.0002	0.0001
	SA	20570	0.0212	0.0213	0.0214	0.0213	0.0214	0.0214	-0.0002	0.0000	0.0000	0.0001	0.0001
	TAS	6584	0.0386	0.0398	0.0391	0.0385	0.0391	0.0391	-0.0012	-0.0007	-0.0013	-0.0008	-0.0008
	ACT	5229	0.0328	0.0343	0.0340	0.0344	0.0333	0.0338	-0.0014	-0.0003	0.0001	-0.0009	-0.0005
5	NT	3501	0.0810	0.0817	0.0812	0.0808	0.0811	0.0812	-0.0007	-0.0005	-0.0009	-0.0006	-0.0005
	Aus	286154	0.0090	0.0090	0.0091	0.0090	0.0090	0.0090	0.0000	0.0001	0.0000	0.0000	0.0000
	NSW	90745	0.0179	0.0179	0.0179	0.0179	0.0179	0.0179	0.0000	0.0000	0.0000	0.0000	0.0000
	VIC	69171	0.0152	0.0153	0.0154	0.0153	0.0153	0.0153	0.0000	0.0001	0.0001	0.0001	0.0001
	QLD	61835	0.0179	0.0179	0.0179	0.0178	0.0178	0.0179	0.0000	0.0000	-0.0001	-0.0001	-0.0001
	WA	31266	0.0239	0.0241	0.0242	0.0240	0.0241	0.0241	-0.0002	0.0001	-0.0001	0.0000	0.0000
	SA	19088	0.0262	0.0262	0.0264	0.0263	0.0263	0.0263	0.0000	0.0003	0.0001	0.0001	0.0001
	TAS	6008	0.0439	0.0430	0.0432	0.0431	0.0435	0.0434	0.0009	0.0002	0.0001	0.0005	0.0004
7	ACT	4786	0.0519	0.0526	0.0514	0.0506	0.0512	0.0513	-0.0006	-0.0012	-0.0020	-0.0014	-0.0013
	NT	3255	0.1184	0.1184	0.1188	0.1182	0.1185	0.1184	0.0000	0.0003	-0.0002	0.0001	-0.0001
	Aus	278343	0.0143	0.0143	0.0143	0.0143	0.0143	0.0143	0.0000	0.0000	0.0000	0.0000	0.0000
	NSW	88648	0.0309	0.0309	0.0309	0.0310	0.0309	0.0309	-0.0001	0.0000	0.0000	0.0000	0.0000
	VIC	68159	0.0251	0.0251	0.0250	0.0250	0.0250	0.0250	0.0000	-0.0001	-0.0001	-0.0001	-0.0001
	QLD	57996	0.0274	0.0274	0.0274	0.0275	0.0274	0.0274	0.0001	0.0000	0.0001	0.0000	0.0001
	WA	30065	0.0447	0.0444	0.0447	0.0445	0.0445	0.0445	0.0003	0.0003	0.0002	0.0001	0.0001
	SA	18974	0.0264	0.0263	0.0263	0.0262	0.0264	0.0263	0.0001	0.0000	-0.0001	0.0001	0.0000
9	TAS	6292	0.0526	0.0517	0.0516	0.0520	0.0522	0.0518	0.0009	-0.0001	0.0003	0.0006	0.0002
	ACT	4886	0.0751	0.0756	0.0755	0.0753	0.0754	0.0751	-0.0004	0.0000	-0.0003	-0.0002	-0.0004
	NT	3323	0.1467	0.1459	0.1476	0.1463	0.1465	0.1470	0.0007	0.0017	0.0003	0.0006	0.0011
	Aus	281602	0.0161	0.0161	0.0160	0.0161	0.0161	0.0161	0.0000	0.0000	0.0000	0.0000	0.0000
	NSW	89624	0.0325	0.0325	0.0325	0.0325	0.0325	0.0325	0.0000	0.0000	0.0000	0.0000	0.0000
	VIC	68446	0.0326	0.0327	0.0326	0.0326	0.0326	0.0326	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
	QLD	59293	0.0284	0.0284	0.0285	0.0284	0.0284	0.0284	0.0000	0.0001	0.0001	0.0000	0.0000
	WA	30256	0.0475	0.0479	0.0477	0.0477	0.0478	0.0477	-0.0003	-0.0002	-0.0002	-0.0001	-0.0002
9	SA	19451	0.0473	0.0473	0.0471	0.0471	0.0474	0.0473	0.0000	-0.0001	-0.0001	0.0001	0.0001
	TAS	6512	0.0582	0.0576	0.0576	0.0577	0.0577	0.0574	0.0005	0.0000	0.0001	0.0001	-0.0002
	ACT	4997	0.0860	0.0873	0.0857	0.0863	0.0865	0.0862	-0.0013	-0.0016	-0.0010	-0.0008	-0.0011
	NT	3023	0.1215	0.1223	0.1222	0.1224	0.1227	0.1225	-0.0008	-0.0001	0.0001	0.0004	0.0002

Table 36 Standard Error Estimates of Standard Deviation from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aus	307549	0.0031	0.0030	0.0035	0.0030	0.0034	0.0032	0.0001	0.0005	0.0001	0.0005	0.0002
	NSW	97404	0.0056	0.0055	0.0053	0.0055	0.0056	0.0054	0.0001	-0.0002	0.0000	0.0001	-0.0001
	VIC	74689	0.0055	0.0067	0.0052	0.0049	0.0056	0.0055	-0.0012	-0.0015	-0.0018	-0.0011	-0.0012
	QLD	65420	0.0057	0.0059	0.0068	0.0056	0.0059	0.0058	-0.0002	0.0009	-0.0003	0.0000	-0.0001
	WA	34152	0.0086	0.0090	0.0105	0.0092	0.0096	0.0093	-0.0004	0.0015	0.0001	0.0006	0.0003
	SA	20570	0.0103	0.0094	0.0092	0.0087	0.0097	0.0097	0.0009	-0.0002	-0.0007	0.0003	0.0002
	TAS	6584	0.0239	0.0241	0.0231	0.0219	0.0240	0.0247	-0.0002	-0.0011	-0.0023	-0.0001	0.0005
	ACT	5229	0.0171	0.0127	0.0153	0.0143	0.0169	0.0160	0.0044	0.0025	0.0016	0.0042	0.0032
	NT	3501	0.0379	0.0370	0.0382	0.0394	0.0412	0.0397	0.0008	0.0012	0.0024	0.0042	0.0027
5	Aus	286154	0.0049	0.0048	0.0049	0.0051	0.0052	0.0053	0.0001	0.0002	0.0004	0.0005	0.0005
	NSW	90745	0.0097	0.0101	0.0097	0.0107	0.0099	0.0102	-0.0004	-0.0004	0.0006	-0.0002	0.0001
	VIC	69171	0.0073	0.0062	0.0074	0.0073	0.0073	0.0069	0.0011	0.0013	0.0011	0.0012	0.0007
	QLD	61835	0.0073	0.0092	0.0077	0.0082	0.0081	0.0082	-0.0019	-0.0015	-0.0010	-0.0010	-0.0010
	WA	31266	0.0127	0.0134	0.0119	0.0125	0.0132	0.0123	-0.0007	-0.0016	-0.0010	-0.0003	-0.0011
	SA	19088	0.0106	0.0138	0.0119	0.0128	0.0124	0.0118	-0.0033	-0.0019	-0.0010	-0.0014	-0.0020
	TAS	6008	0.0200	0.0254	0.0260	0.0229	0.0220	0.0244	-0.0054	0.0006	-0.0024	-0.0034	-0.0010
	ACT	4786	0.0200	0.0236	0.0293	0.0221	0.0227	0.0235	-0.0036	0.0057	-0.0015	-0.0009	-0.0001
	NT	3255	0.0512	0.0548	0.0682	0.0596	0.0592	0.0626	-0.0036	0.0134	0.0047	0.0044	0.0078
7	Aus	278343	0.0112	0.0111	0.0110	0.0112	0.0111	0.0111	0.0001	-0.0001	0.0000	0.0000	0.0000
	NSW	88648	0.0263	0.0264	0.0262	0.0265	0.0264	0.0264	-0.0001	-0.0002	0.0001	-0.0001	0.0000
	VIC	68159	0.0121	0.0122	0.0122	0.0119	0.0121	0.0121	0.0000	0.0000	-0.0003	0.0000	-0.0001
	QLD	57996	0.0136	0.0135	0.0131	0.0133	0.0133	0.0133	0.0001	-0.0004	-0.0002	-0.0002	-0.0002
	WA	30065	0.0270	0.0271	0.0275	0.0270	0.0269	0.0268	-0.0001	0.0004	-0.0001	-0.0002	-0.0003
	SA	18974	0.0150	0.0158	0.0155	0.0149	0.0151	0.0148	-0.0008	-0.0003	-0.0010	-0.0007	-0.0010
	TAS	6292	0.0219	0.0229	0.0211	0.0219	0.0225	0.0226	-0.0010	-0.0018	-0.0010	-0.0004	-0.0003
	ACT	4886	0.0326	0.0263	0.0280	0.0292	0.0274	0.0276	0.0063	0.0017	0.0029	0.0012	0.0013
	NT	3323	0.0745	0.0773	0.0737	0.0736	0.0747	0.0730	-0.0028	-0.0036	-0.0037	-0.0026	-0.0043
9	Aus	281602	0.0128	0.0124	0.0124	0.0125	0.0124	0.0124	0.0004	0.0000	0.0001	0.0000	0.0000
	NSW	89624	0.0271	0.0265	0.0265	0.0266	0.0267	0.0267	0.0006	0.0000	0.0001	0.0001	0.0002
	VIC	68446	0.0234	0.0233	0.0233	0.0230	0.0230	0.0232	0.0001	0.0000	-0.0003	-0.0003	-0.0001
	QLD	59293	0.0157	0.0152	0.0152	0.0158	0.0151	0.0153	0.0005	0.0000	0.0006	-0.0001	0.0001
	WA	30256	0.0262	0.0258	0.0264	0.0258	0.0261	0.0261	0.0004	0.0006	0.0000	0.0003	0.0003
	SA	19451	0.0246	0.0259	0.0244	0.0246	0.0253	0.0251	-0.0014	-0.0015	-0.0013	-0.0006	-0.0008
	TAS	6512	0.0235	0.0234	0.0250	0.0248	0.0240	0.0237	0.0001	0.0016	0.0013	0.0006	0.0003
	ACT	4997	0.0370	0.0414	0.0373	0.0373	0.0351	0.0362	-0.0043	-0.0041	-0.0041	-0.0063	-0.0052
	NT	3023	0.0548	0.0601	0.0578	0.0569	0.0591	0.0569	-0.0052	-0.0023	-0.0032	-0.0010	-0.0031

Table 37 Standard Error Estimates of Percentage in the Top Band from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aus	307549	0.2213	0.2181	0.2170	0.2176	0.2163	0.2182	0.0032	-0.0011	-0.0005	-0.0018	0.0001
	NSW	97404	0.4399	0.4253	0.4301	0.4260	0.4252	0.4258	0.0146	0.0048	0.0007	-0.0001	0.0005
	VIC	74689	0.4542	0.4411	0.4416	0.4326	0.4372	0.4377	0.0131	0.0005	-0.0085	-0.0039	-0.0034
	QLD	65420	0.4253	0.4226	0.4125	0.4204	0.4232	0.4186	0.0027	-0.0100	-0.0022	0.0006	-0.0039
	WA	34152	0.5553	0.5786	0.5734	0.5719	0.5752	0.5679	-0.0233	-0.0052	-0.0067	-0.0034	-0.0106
	SA	20570	0.5402	0.5474	0.5539	0.5617	0.5776	0.5719	-0.0072	0.0064	0.0142	0.0301	0.0245
	TAS	6584	1.1889	1.3039	1.2882	1.3057	1.2788	1.2808	-0.1150	-0.0158	0.0018	-0.0251	-0.0231
	ACT	5229	1.3859	1.2230	1.2980	1.2657	1.2399	1.2433	0.1628	0.0749	0.0427	0.0169	0.0203
	NT	3501	0.9388	1.0843	0.9723	0.9626	1.0187	0.9918	-0.1454	-0.1119	-0.1216	-0.0655	-0.0925
5	Aus	286154	0.2500	0.2444	0.2444	0.2439	0.2466	0.2460	0.0056	0.0001	-0.0005	0.0022	0.0016
	NSW	90745	0.5093	0.5066	0.5039	0.5062	0.5057	0.5072	0.0027	-0.0026	-0.0004	-0.0009	0.0006
	VIC	69171	0.4589	0.4673	0.4755	0.4654	0.4685	0.4670	-0.0084	0.0083	-0.0019	0.0013	-0.0003
	QLD	61835	0.4513	0.4563	0.4539	0.4504	0.4490	0.4488	-0.0051	-0.0024	-0.0059	-0.0074	-0.0076
	WA	31266	0.6104	0.6153	0.6095	0.6177	0.6099	0.6141	-0.0049	-0.0058	0.0023	-0.0054	-0.0012
	SA	19088	0.5968	0.5979	0.6176	0.6028	0.6151	0.6056	-0.0011	0.0197	0.0049	0.0172	0.0077
	TAS	6008	1.1304	1.1474	1.1490	1.1538	1.1892	1.1744	-0.0170	0.0016	0.0064	0.0417	0.0270
	ACT	4786	1.6606	1.5939	1.6394	1.5827	1.6054	1.6084	0.0667	0.0456	-0.0112	0.0115	0.0146
	NT	3255	1.2754	1.2541	1.2331	1.2830	1.2434	1.2760	0.0214	-0.0210	0.0289	-0.0106	0.0219
7	Aus	278343	0.4050	0.4057	0.4047	0.4041	0.4042	0.4052	-0.0007	-0.0009	-0.0015	-0.0015	-0.0005
	NSW	88648	0.8595	0.8652	0.8655	0.8634	0.8640	0.8628	-0.0057	0.0003	-0.0019	-0.0012	-0.0024
	VIC	68159	0.7690	0.7750	0.7686	0.7674	0.7688	0.7700	-0.0060	-0.0064	-0.0076	-0.0062	-0.0050
	QLD	57996	0.7670	0.7641	0.7654	0.7627	0.7625	0.7677	0.0029	0.0014	-0.0014	-0.0016	0.0036
	WA	30065	1.2458	1.2262	1.2339	1.2296	1.2350	1.2285	0.0196	0.0077	0.0034	0.0088	0.0023
	SA	18974	0.7370	0.7208	0.7274	0.7357	0.7362	0.7345	0.0162	0.0066	0.0149	0.0154	0.0137
	TAS	6292	1.2954	1.2623	1.2536	1.2851	1.2650	1.2497	0.0330	-0.0087	0.0227	0.0027	-0.0126
	ACT	4886	2.3630	2.3477	2.3573	2.3197	2.3404	2.3198	0.0152	0.0096	-0.0281	-0.0073	-0.0279
	NT	3323	1.6146	1.6982	1.7607	1.7382	1.7385	1.7569	-0.0835	0.0625	0.0400	0.0404	0.0587
9	Aus	281602	0.4471	0.4468	0.4478	0.4468	0.4467	0.4471	0.0004	0.0010	0.0000	-0.0001	0.0003
	NSW	89624	0.8812	0.8889	0.8822	0.8831	0.8838	0.8833	-0.0077	-0.0067	-0.0058	-0.0051	-0.0056
	VIC	68446	0.9643	0.9564	0.9547	0.9565	0.9549	0.9557	0.0079	-0.0017	0.0001	-0.0015	-0.0007
	QLD	59293	0.7788	0.7755	0.7784	0.7777	0.7748	0.7745	0.0033	0.0029	0.0022	-0.0007	-0.0010
	WA	30256	1.3877	1.3897	1.4041	1.3924	1.3956	1.3966	-0.0020	0.0144	0.0026	0.0059	0.0069
	SA	19451	1.1755	1.1700	1.1526	1.1514	1.1621	1.1627	0.0055	-0.0175	-0.0186	-0.0080	-0.0073
	TAS	6512	1.4706	1.4262	1.4455	1.4344	1.4503	1.4323	0.0445	0.0194	0.0083	0.0242	0.0062
	ACT	4997	2.5237	2.6007	2.5380	2.4794	2.5306	2.4984	-0.0770	-0.0627	-0.1213	-0.0701	-0.1023
	NT	3023	1.6325	1.5543	1.5292	1.5457	1.5057	1.5367	0.0782	-0.0251	-0.0086	-0.0485	-0.0176

Table 38 Standard Error Estimates of 20th Percentile from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aus	307549	0.0820	0.0821	0.0822	0.0821	0.0821	0.0819	-0.0001	0.0002	0.0000	0.0001	-0.0001
	NSW	97404	0.1127	0.1119	0.1118	0.1112	0.1112	0.1109	0.0009	-0.0001	-0.0007	-0.0006	-0.0009
	VIC	74689	0.1101	0.1080	0.1101	0.1091	0.1093	0.1095	0.0021	0.0022	0.0011	0.0014	0.0015
	QLD	65420	0.1181	0.1177	0.1193	0.1189	0.1199	0.1196	0.0004	0.0016	0.0012	0.0022	0.0018
	WA	34152	0.1414	0.1396	0.1397	0.1387	0.1397	0.1398	0.0018	0.0002	-0.0009	0.0001	0.0002
	SA	20570	0.1497	0.1473	0.1515	0.1497	0.1485	0.1500	0.0025	0.0042	0.0024	0.0012	0.0027
	TAS	6584	0.1897	0.1940	0.1936	0.1895	0.1915	0.1914	-0.0043	-0.0003	-0.0044	-0.0025	-0.0026
	ACT	5229	0.1821	0.1948	0.1921	0.1895	0.1891	0.1888	-0.0128	-0.0027	-0.0053	-0.0058	-0.0060
	NT	3501	0.3314	0.3254	0.3302	0.3287	0.3290	0.3280	0.0060	0.0048	0.0034	0.0037	0.0026
5	Aus	286154	0.0899	0.0898	0.0894	0.0895	0.0895	0.0895	0.0001	-0.0005	-0.0003	-0.0003	-0.0004
	NSW	90745	0.1206	0.1223	0.1203	0.1208	0.1209	0.1213	-0.0017	-0.0020	-0.0015	-0.0014	-0.0010
	VIC	69171	0.1184	0.1185	0.1176	0.1190	0.1172	0.1185	0.0000	-0.0009	0.0005	-0.0013	0.0000
	QLD	61835	0.1305	0.1322	0.1312	0.1324	0.1319	0.1318	-0.0017	-0.0011	0.0002	-0.0003	-0.0004
	WA	31266	0.1512	0.1501	0.1521	0.1507	0.1514	0.1502	0.0012	0.0020	0.0007	0.0014	0.0001
	SA	19088	0.1611	0.1573	0.1590	0.1587	0.1591	0.1591	0.0038	0.0017	0.0014	0.0018	0.0019
	TAS	6008	0.2025	0.1955	0.1932	0.1962	0.1976	0.1966	0.0070	-0.0024	0.0007	0.0021	0.0011
	ACT	4786	0.2233	0.2265	0.2206	0.2224	0.2205	0.2192	-0.0032	-0.0059	-0.0041	-0.0060	-0.0074
	NT	3255	0.3726	0.3742	0.3823	0.3826	0.3849	0.3836	-0.0017	0.0080	0.0083	0.0106	0.0094
7	Aus	278343	0.0993	0.1000	0.1002	0.0994	0.0998	0.0994	-0.0007	0.0002	-0.0006	-0.0002	-0.0005
	NSW	88648	0.1335	0.1348	0.1345	0.1339	0.1343	0.1338	-0.0013	-0.0002	-0.0009	-0.0005	-0.0009
	VIC	68159	0.1370	0.1371	0.1373	0.1372	0.1376	0.1379	-0.0001	0.0001	0.0001	0.0004	0.0007
	QLD	57996	0.1497	0.1499	0.1507	0.1503	0.1504	0.1492	-0.0002	0.0008	0.0004	0.0005	-0.0007
	WA	30065	0.1875	0.1860	0.1844	0.1853	0.1836	0.1851	0.0016	-0.0015	-0.0007	-0.0023	-0.0009
	SA	18974	0.1421	0.1448	0.1469	0.1442	0.1459	0.1457	-0.0027	0.0021	-0.0006	0.0011	0.0009
	TAS	6292	0.2118	0.2119	0.2109	0.2129	0.2126	0.2110	-0.0001	-0.0010	0.0010	0.0007	-0.0009
	ACT	4886	0.2494	0.2463	0.2453	0.2439	0.2437	0.2436	0.0031	-0.0010	-0.0024	-0.0026	-0.0026
	NT	3323	0.4494	0.4431	0.4443	0.4353	0.4374	0.4362	0.0062	0.0012	-0.0079	-0.0057	-0.0069
9	Aus	281602	0.1027	0.1033	0.1033	0.1035	0.1038	0.1032	-0.0006	-0.0001	0.0002	0.0005	-0.0001
	NSW	89624	0.1396	0.1382	0.1384	0.1390	0.1394	0.1392	0.0014	0.0001	0.0008	0.0012	0.0009
	VIC	68446	0.1460	0.1455	0.1466	0.1470	0.1460	0.1469	0.0005	0.0011	0.0015	0.0005	0.0014
	QLD	59293	0.1463	0.1466	0.1460	0.1467	0.1467	0.1463	-0.0003	-0.0006	0.0001	0.0000	-0.0003
	WA	30256	0.1912	0.1920	0.1930	0.1942	0.1931	0.1928	-0.0008	0.0009	0.0022	0.0010	0.0007
	SA	19451	0.1922	0.1942	0.1935	0.1911	0.1934	0.1932	-0.0020	-0.0007	-0.0031	-0.0008	-0.0010
	TAS	6512	0.2142	0.2123	0.2128	0.2137	0.2113	0.2117	0.0019	0.0005	0.0014	-0.0010	-0.0006
	ACT	4997	0.2704	0.2800	0.2726	0.2763	0.2764	0.2752	-0.0095	-0.0074	-0.0037	-0.0036	-0.0048
	NT	3023	0.3664	0.3828	0.3788	0.3818	0.3812	0.3845	-0.0165	-0.0041	-0.0011	-0.0016	0.0017

Table 39 Standard Error Estimates of 80th Percentile from All Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aus	307549	0.0882	0.0887	0.0887	0.0884	0.0890	0.0888	-0.0004	0.0000	-0.0003	0.0003	0.0002
	NSW	97404	0.1236	0.1243	0.1235	0.1244	0.1236	0.1235	-0.0007	-0.0008	0.0001	-0.0007	-0.0008
	VIC	74689	0.1184	0.1181	0.1196	0.1194	0.1197	0.1199	0.0003	0.0015	0.0013	0.0017	0.0018
	QLD	65420	0.1259	0.1245	0.1248	0.1248	0.1237	0.1246	0.0014	0.0004	0.0004	-0.0007	0.0001
	WA	34152	0.1492	0.1483	0.1487	0.1480	0.1485	0.1477	0.0009	0.0004	-0.0003	0.0003	-0.0006
	SA	20570	0.1547	0.1534	0.1522	0.1523	0.1520	0.1523	0.0013	-0.0012	-0.0011	-0.0014	-0.0010
	TAS	6584	0.2097	0.2191	0.2173	0.2168	0.2161	0.2162	-0.0094	-0.0018	-0.0023	-0.0030	-0.0029
	ACT	5229	0.1836	0.1950	0.1990	0.1961	0.1963	0.1963	-0.0114	0.0040	0.0012	0.0013	0.0013
	NT	3501	0.2657	0.2705	0.2650	0.2681	0.2651	0.2653	-0.0048	-0.0055	-0.0024	-0.0054	-0.0052
5	Aus	286154	0.1052	0.1055	0.1054	0.1049	0.1048	0.1051	-0.0003	-0.0001	-0.0006	-0.0007	-0.0004
	NSW	90745	0.1550	0.1539	0.1558	0.1552	0.1545	0.1545	0.0010	0.0019	0.0013	0.0006	0.0006
	VIC	69171	0.1377	0.1367	0.1369	0.1383	0.1365	0.1374	0.0010	0.0003	0.0016	-0.0001	0.0007
	QLD	61835	0.1425	0.1414	0.1433	0.1428	0.1425	0.1429	0.0011	0.0019	0.0013	0.0011	0.0015
	WA	31266	0.1688	0.1712	0.1680	0.1711	0.1692	0.1696	-0.0023	-0.0032	-0.0001	-0.0020	-0.0016
	SA	19088	0.1810	0.1784	0.1780	0.1759	0.1786	0.1774	0.0027	-0.0004	-0.0024	0.0002	-0.0009
	TAS	6008	0.2284	0.2385	0.2317	0.2361	0.2376	0.2366	-0.0101	-0.0068	-0.0024	-0.0009	-0.0019
	ACT	4786	0.2568	0.2518	0.2473	0.2504	0.2501	0.2492	0.0050	-0.0044	-0.0014	-0.0017	-0.0026
	NT	3255	0.3348	0.3309	0.3242	0.3319	0.3253	0.3259	0.0040	-0.0066	0.0010	-0.0056	-0.0049
7	Aus	278343	0.1380	0.1384	0.1378	0.1381	0.1382	0.1379	-0.0004	-0.0007	-0.0004	-0.0002	-0.0006
	NSW	88648	0.2131	0.2120	0.2130	0.2118	0.2117	0.2123	0.0012	0.0011	-0.0002	-0.0003	0.0003
	VIC	68159	0.1841	0.1882	0.1858	0.1851	0.1865	0.1860	-0.0041	-0.0024	-0.0031	-0.0016	-0.0022
	QLD	57996	0.1839	0.1828	0.1859	0.1837	0.1840	0.1846	0.0011	0.0031	0.0009	0.0012	0.0018
	WA	30065	0.2389	0.2366	0.2398	0.2384	0.2396	0.2390	0.0022	0.0031	0.0017	0.0030	0.0023
	SA	18974	0.1878	0.1892	0.1855	0.1873	0.1864	0.1861	-0.0014	-0.0037	-0.0019	-0.0028	-0.0031
	TAS	6292	0.2579	0.2579	0.2556	0.2594	0.2609	0.2584	0.0001	-0.0023	0.0015	0.0030	0.0006
	ACT	4886	0.3237	0.3263	0.3199	0.3211	0.3222	0.3206	-0.0027	-0.0064	-0.0052	-0.0041	-0.0058
	NT	3323	0.3904	0.3881	0.3929	0.3909	0.3888	0.3894	0.0023	0.0048	0.0028	0.0007	0.0013
9	Aus	281602	0.1457	0.1463	0.1463	0.1456	0.1459	0.1461	-0.0006	0.0000	-0.0007	-0.0004	-0.0002
	NSW	89624	0.2154	0.2142	0.2143	0.2151	0.2156	0.2149	0.0012	0.0002	0.0009	0.0015	0.0008
	VIC	68446	0.2101	0.2091	0.2122	0.2113	0.2110	0.2107	0.0010	0.0031	0.0023	0.0020	0.0016
	QLD	59293	0.1911	0.1904	0.1912	0.1911	0.1916	0.1912	0.0007	0.0007	0.0006	0.0012	0.0008
	WA	30256	0.2471	0.2519	0.2497	0.2488	0.2492	0.2492	-0.0048	-0.0021	-0.0030	-0.0027	-0.0026
	SA	19451	0.2436	0.2438	0.2438	0.2434	0.2423	0.2428	-0.0002	0.0001	-0.0004	-0.0015	-0.0010
	TAS	6512	0.2850	0.2745	0.2797	0.2770	0.2778	0.2771	0.0105	0.0052	0.0025	0.0033	0.0026
	ACT	4997	0.3247	0.3289	0.3192	0.3206	0.3266	0.3233	-0.0042	-0.0097	-0.0083	-0.0023	-0.0056
	NT	3023	0.3769	0.3714	0.3734	0.3738	0.3723	0.3735	0.0054	0.0020	0.0024	0.0008	0.0021

Table 40 Estimates of Mean from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	16700	-0.6833	-0.6852	-0.6804	-0.6847	-0.6820	-0.6830	-0.6838	-0.0030	-0.0048	-0.0043	-0.0016	-0.0026	-0.0034
	NSW	5370	-0.5222	-0.5231	-0.5143	-0.5188	-0.5178	-0.5163	-0.5201	-0.0080	-0.0089	-0.0045	-0.0035	-0.0020	-0.0058
	QLD	4723	-0.5745	-0.5872	-0.5805	-0.5883	-0.5821	-0.5845	-0.5823	0.0059	-0.0067	-0.0078	-0.0016	-0.0041	-0.0018
	WA	2358	-0.9754	-0.9764	-0.9797	-0.9734	-0.9785	-0.9781	-0.9801	0.0043	0.0032	0.0063	0.0012	0.0015	-0.0004
	NT	1478	-1.4819	-1.4602	-1.4626	-1.4741	-1.4608	-1.4689	-1.4692	-0.0193	0.0025	-0.0115	0.0018	-0.0063	-0.0066
	VIC	1164	-0.2597	-0.2700	-0.2642	-0.2631	-0.2684	-0.2640	-0.2655	0.0045	-0.0058	0.0011	-0.0042	0.0003	-0.0012
	SA	899	-0.9159	-0.9132	-0.9000	-0.9086	-0.8993	-0.9054	-0.9031	-0.0159	-0.0131	-0.0086	0.0007	-0.0053	-0.0031
	TAS	551	-0.3454	-0.3230	-0.3302	-0.3261	-0.3197	-0.3225	-0.3238	-0.0152	0.0072	0.0041	0.0105	0.0077	0.0064
5	ACT	157	-0.3705	-0.3634	-0.3790	-0.4131	-0.4301	-0.4247	-0.4209	0.0085	0.0155	-0.0342	-0.0511	-0.0458	-0.0419
	Aus	15507	-0.7951	-0.7967	-0.7972	-0.7965	-0.7980	-0.7971	-0.7973	0.0020	0.0005	0.0007	-0.0008	0.0001	-0.0001
	NSW	4899	-0.5723	-0.5846	-0.5775	-0.5773	-0.5808	-0.5783	-0.5781	0.0052	-0.0071	0.0002	-0.0033	-0.0007	-0.0006
	QLD	4508	-0.6938	-0.6877	-0.6858	-0.6860	-0.6861	-0.6868	-0.6871	-0.0080	-0.0019	-0.0002	-0.0003	-0.0010	-0.0013
	WA	2167	-1.0890	-1.0817	-1.0950	-1.0870	-1.0911	-1.0920	-1.0900	0.0061	0.0133	0.0080	0.0039	0.0030	0.0050
	NT	1440	-1.7466	-1.7639	-1.7705	-1.7779	-1.7744	-1.7753	-1.7716	0.0239	0.0067	-0.0074	-0.0039	-0.0048	-0.0011
	VIC	1053	-0.3920	-0.3707	-0.3864	-0.3856	-0.3840	-0.3793	-0.3857	-0.0056	0.0157	0.0009	0.0024	0.0072	0.0007
	SA	821	-1.0057	-1.0019	-0.9951	-0.9961	-0.9917	-0.9971	-0.9978	-0.0106	-0.0068	-0.0010	0.0033	-0.0020	-0.0027
7	TAS	479	-0.3686	-0.3853	-0.3939	-0.3860	-0.3999	-0.3908	-0.3964	0.0253	0.0086	0.0079	-0.0060	0.0032	-0.0024
	ACT	140	-0.5232	-0.5123	-0.4963	-0.4966	-0.4963	-0.4760	-0.4945	-0.0270	-0.0160	-0.0004	0.0000	0.0203	0.0017
	Aus	15083	-1.0977	-1.0939	-1.0962	-1.0954	-1.0969	-1.0967	-1.0961	-0.0016	0.0022	0.0008	-0.0008	-0.0005	0.0001
	NSW	4829	-0.9725	-0.9752	-0.9754	-0.9774	-0.9803	-0.9786	-0.9778	0.0030	0.0002	-0.0020	-0.0049	-0.0032	-0.0023
	QLD	4249	-0.9674	-0.9493	-0.9573	-0.9566	-0.9558	-0.9582	-0.9549	-0.0101	0.0080	0.0007	0.0015	-0.0009	0.0024
	WA	1989	-1.2644	-1.2617	-1.2650	-1.2611	-1.2732	-1.2642	-1.2687	0.0006	0.0033	0.0040	-0.0081	0.0008	-0.0036
	NT	1453	-1.9601	-1.9804	-1.9737	-1.9691	-1.9734	-1.9737	-1.9733	0.0136	-0.0067	0.0046	0.0004	0.0000	0.0005
	VIC	1123	-0.8404	-0.8320	-0.8401	-0.8342	-0.8298	-0.8310	-0.8342	-0.0003	0.0081	0.0059	0.0102	0.0091	0.0059
9	SA	794	-1.1225	-1.1165	-1.1056	-1.1121	-1.1051	-1.1087	-1.1090	-0.0169	-0.0109	-0.0065	0.0005	-0.0031	-0.0034
	TAS	518	-0.8552	-0.8424	-0.8419	-0.8407	-0.8307	-0.8396	-0.8371	-0.0133	-0.0005	0.0012	0.0112	0.0024	0.0049
	ACT	128	-0.7770	-0.8005	-0.8133	-0.7953	-0.7821	-0.8050	-0.7957	0.0363	0.0127	0.0179	0.0311	0.0083	0.0175
	Aus	14358	-1.3086	-1.3054	-1.3076	-1.3077	-1.3075	-1.3078	-1.3066	-0.0010	0.0022	-0.0001	0.0002	-0.0002	0.0010
	NSW	4624	-1.2459	-1.2486	-1.2422	-1.2423	-1.2434	-1.2410	-1.2407	-0.0037	-0.0064	0.0000	-0.0012	0.0013	0.0015
	QLD	4043	-1.2052	-1.2032	-1.2075	-1.2043	-1.2043	-1.2038	-1.2057	0.0022	0.0043	0.0031	0.0032	0.0037	0.0017
	WA	1912	-1.3974	-1.3840	-1.3905	-1.3896	-1.3913	-1.3895	-1.3848	-0.0069	0.0065	0.0009	-0.0008	0.0010	0.0057
	NT	1292	-2.0091	-1.9967	-2.0229	-2.0194	-2.0199	-2.0203	-2.0159	0.0138	0.0262	0.0035	0.0030	0.0026	0.0070
9	VIC	1025	-1.0576	-1.0323	-1.0387	-1.0476	-1.0309	-1.0444	-1.0380	-0.0188	0.0065	-0.0088	0.0078	-0.0056	0.0008
	SA	820	-1.3803	-1.3845	-1.3669	-1.3861	-1.3844	-1.3897	-1.3890	-0.0134	-0.0176	-0.0191	-0.0175	-0.0227	-0.0221
	TAS	514	-1.0289	-1.0709	-1.0746	-1.0687	-1.0749	-1.0815	-1.0740	0.0456	0.0037	0.0058	-0.0004	-0.0070	0.0005
	ACT	128	-1.0485	-0.9908	-1.0135	-1.0092	-1.0184	-1.0247	-1.0363	-0.0350	0.0227	0.0044	-0.0049	-0.0112	-0.0228

Table 41 Estimates of Standard Deviation from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	16700	0.9574	0.9492	0.9497	0.9529	0.9506	0.9522	0.9512	0.0077	-0.0005	0.0032	0.0009	0.0025	0.0015
	NSW	5370	0.8970	0.8897	0.8925	0.8936	0.8913	0.8946	0.8923	0.0044	-0.0028	0.0010	-0.0012	0.0020	-0.0003
	QLD	4723	0.8759	0.8819	0.8768	0.8783	0.8810	0.8784	0.8785	-0.0009	0.0051	0.0015	0.0042	0.0016	0.0018
	WA	2358	0.9333	0.9021	0.9122	0.9169	0.9097	0.9130	0.9130	0.0211	-0.0101	0.0046	-0.0026	0.0007	0.0007
	NT	1478	0.9776	0.9691	0.9615	0.9641	0.9577	0.9616	0.9614	0.0160	0.0076	0.0026	-0.0038	0.0001	-0.0002
	VIC	1164	0.8953	0.8826	0.8855	0.8883	0.8946	0.8938	0.8873	0.0098	-0.0028	0.0028	0.0092	0.0084	0.0019
	SA	899	0.9032	0.9013	0.8783	0.8926	0.8942	0.8867	0.8907	0.0249	0.0230	0.0143	0.0159	0.0084	0.0124
	TAS	551	0.9010	0.8976	0.9100	0.9188	0.9120	0.9165	0.9136	-0.0089	-0.0124	0.0088	0.0020	0.0065	0.0036
	ACT	157	0.8584	0.9131	0.9076	0.9381	0.9100	0.9123	0.9212	-0.0492	0.0055	0.0305	0.0024	0.0047	0.0136
5	Aus	15507	1.1087	1.1052	1.1088	1.1081	1.1058	1.1078	1.1080	-0.0001	-0.0036	-0.0006	-0.0029	-0.0010	-0.0008
	NSW	4899	1.0316	1.0316	1.0288	1.0289	1.0324	1.0304	1.0320	0.0028	0.0028	0.0001	0.0036	0.0016	0.0032
	QLD	4508	1.0363	1.0356	1.0311	1.0318	1.0265	1.0298	1.0313	0.0052	0.0045	0.0006	-0.0046	-0.0013	0.0002
	WA	2167	1.0812	1.0565	1.0712	1.0622	1.0625	1.0608	1.0626	0.0100	-0.0148	-0.0090	-0.0087	-0.0104	-0.0086
	NT	1440	1.1138	1.1403	1.1485	1.1410	1.1383	1.1493	1.1434	-0.0347	-0.0082	-0.0076	-0.0102	0.0007	-0.0051
	VIC	1053	1.0238	0.9872	1.0020	1.0071	0.9948	0.9953	1.0008	0.0218	-0.0147	0.0051	-0.0072	-0.0066	-0.0012
	SA	821	1.0098	1.0011	1.0142	1.0147	1.0156	1.0151	1.0161	-0.0044	-0.0130	0.0006	0.0015	0.0009	0.0020
	TAS	479	1.0317	1.0179	1.0197	1.0304	1.0173	1.0216	1.0190	0.0120	-0.0018	0.0107	-0.0024	0.0019	-0.0007
	ACT	140	1.0155	0.9624	0.9968	0.9971	0.9895	1.0011	0.9967	0.0186	-0.0345	0.0002	-0.0074	0.0043	-0.0002
7	Aus	15083	0.9419	0.9480	0.9458	0.9484	0.9469	0.9483	0.9477	-0.0039	0.0022	0.0026	0.0011	0.0025	0.0019
	NSW	4829	0.8697	0.8792	0.8811	0.8813	0.8813	0.8821	0.8803	-0.0114	-0.0019	0.0002	0.0002	0.0010	-0.0008
	QLD	4249	0.8813	0.8815	0.8748	0.8823	0.8797	0.8822	0.8800	0.0065	0.0068	0.0076	0.0049	0.0075	0.0053
	WA	1989	0.8391	0.8462	0.8404	0.8461	0.8478	0.8427	0.8458	-0.0013	0.0058	0.0056	0.0073	0.0023	0.0054
	NT	1453	1.0856	1.0936	1.0916	1.0987	1.0880	1.0884	1.0914	-0.0060	0.0021	0.0071	-0.0036	-0.0031	-0.0001
	VIC	1123	0.8824	0.8832	0.8852	0.8781	0.8808	0.8916	0.8818	-0.0028	-0.0020	-0.0071	-0.0044	0.0064	-0.0034
	SA	794	0.8104	0.7971	0.8093	0.8097	0.7969	0.8059	0.8109	0.0011	-0.0123	0.0003	-0.0124	-0.0034	0.0016
	TAS	518	0.8826	0.8704	0.8765	0.8749	0.8625	0.8708	0.8714	0.0061	-0.0061	-0.0016	-0.0140	-0.0057	-0.0051
	ACT	128	0.9795	0.9576	0.9644	0.9619	0.9801	0.9816	0.9817	0.0151	-0.0068	-0.0026	0.0157	0.0172	0.0173
9	Aus	14358	0.9502	0.9504	0.9517	0.9513	0.9489	0.9499	0.9495	-0.0015	-0.0012	-0.0003	-0.0028	-0.0018	-0.0022
	NSW	4624	0.9242	0.9401	0.9299	0.9349	0.9317	0.9334	0.9337	-0.0057	0.0102	0.0049	0.0018	0.0034	0.0038
	QLD	4043	0.8851	0.8874	0.8914	0.8826	0.8852	0.8820	0.8834	-0.0063	-0.0039	-0.0088	-0.0061	-0.0093	-0.0079
	WA	1912	0.9387	0.9170	0.9296	0.9359	0.9239	0.9309	0.9272	0.0091	-0.0126	0.0063	-0.0057	0.0013	-0.0024
	NT	1292	0.9430	0.9578	0.9576	0.9513	0.9501	0.9541	0.9546	-0.0146	0.0002	-0.0063	-0.0075	-0.0034	-0.0030
	VIC	1025	0.9684	0.9595	0.9606	0.9595	0.9614	0.9664	0.9603	0.0078	-0.0010	-0.0010	0.0009	0.0058	-0.0002
	SA	820	0.9352	0.8991	0.9066	0.9055	0.8917	0.8883	0.8924	0.0287	-0.0074	-0.0011	-0.0148	-0.0182	-0.0142
	TAS	514	0.8706	0.8358	0.8500	0.8516	0.8539	0.8570	0.8539	0.0206	-0.0142	0.0016	0.0039	0.0070	0.0039
	ACT	128	1.0054	1.0629	0.9851	1.0255	1.0095	1.0053	1.0220	0.0204	0.0778	0.0404	0.0244	0.0203	0.0369

Table 42 Estimates of Percentage in the Top Band from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	16700	3.8925	3.8905	3.9491	3.9546	3.9520	3.9343	3.9257	-0.0566	-0.0586	0.0055	0.0029	-0.0148	-0.0234
	NSW	5370	4.7674	4.7419	4.9733	4.8818	4.8373	4.9113	4.8809	-0.2059	-0.2314	-0.0915	-0.1360	-0.0620	-0.0924
	QLD	4723	3.7576	3.9024	3.8923	3.8141	3.9241	3.7576	3.8497	-0.1347	0.0101	-0.0782	0.0319	-0.1347	-0.0426
	WA	2358	2.0137	1.6281	1.5681	1.6924	1.6538	1.6410	1.6356	0.4456	0.0600	0.1243	0.0857	0.0728	0.0674
	NT	1478	0.5552	0.5089	0.4858	0.5760	0.5552	0.5968	0.5144	0.0694	0.0231	0.0902	0.0694	0.1110	0.0287
	VIC	1164	7.6852	8.0864	8.5556	8.4444	8.4198	8.5046	8.1683	-0.8704	-0.4691	-0.1111	-0.1358	-0.0509	-0.3873
	SA	899	2.7714	1.8476	1.6397	2.0323	1.9323	1.8649	1.9856	1.1316	0.2079	0.3926	0.2925	0.2252	0.3459
	TAS	551	7.5646	8.1796	7.7860	8.1550	8.3518	8.4041	8.2103	-0.2214	0.3936	0.3690	0.5658	0.6181	0.4244
5	ACT	157	4.6358	7.5055	6.4901	7.0861	5.8278	5.8278	6.4210	-1.8543	1.0155	0.5960	-0.6623	-0.6623	-0.0691
	Aus	15507	3.4274	3.1440	3.3039	3.2375	3.2499	3.2282	3.2253	0.1235	-0.1599	-0.0664	-0.0540	-0.0757	-0.0786
	NSW	4899	4.4707	4.1614	4.3568	4.2050	4.2950	4.2514	4.2960	0.1139	-0.1954	-0.1518	-0.0619	-0.1054	-0.0608
	QLD	4508	3.3394	3.2712	3.3667	3.3735	3.3348	3.3258	3.3196	-0.0273	-0.0954	0.0068	-0.0318	-0.0409	-0.0470
	WA	2167	1.9718	1.5806	1.7746	1.7324	1.7778	1.6784	1.6493	0.1972	-0.1941	-0.0423	0.0031	-0.0962	-0.1253
	NT	1440	0.4252	0.2599	0.3685	0.2268	0.2882	0.2835	0.2758	0.0567	-0.1087	-0.1417	-0.0803	-0.0850	-0.0927
	VIC	1053	5.7436	5.5043	5.3128	5.4359	5.2991	5.3641	5.3311	0.4308	0.1915	0.1231	-0.0137	0.0513	0.0183
	SA	821	1.7789	1.3554	1.8551	1.6900	1.6518	1.7027	1.6657	-0.0762	-0.4998	-0.1652	-0.2033	-0.1525	-0.1895
7	TAS	479	7.0064	5.5202	5.9448	5.7325	5.7183	5.6369	5.5525	1.0616	-0.4246	-0.2123	-0.2265	-0.3079	-0.3923
	ACT	140	4.3796	3.1630	3.7956	4.4526	3.8443	4.3796	3.9829	0.5839	-0.6326	0.6569	0.0487	0.5839	0.1872
	Aus	15083	2.7211	2.8143	2.7743	2.8985	2.8316	2.8398	2.8263	-0.0532	0.0400	0.1241	0.0573	0.0655	0.0519
	NSW	4829	3.0910	3.3042	3.2999	3.3234	3.2573	3.2850	3.2430	-0.2089	0.0043	0.0234	-0.0426	-0.0149	-0.0569
	QLD	4249	3.2914	3.6141	3.4221	3.6108	3.5528	3.5431	3.5308	-0.1307	0.1920	0.1888	0.1307	0.1210	0.1087
	WA	1989	1.1711	1.1711	1.1914	1.1914	1.2322	1.1558	1.1080	-0.0204	-0.0204	0.0000	0.0407	-0.0356	-0.0835
	NT	1453	0.4958	0.5194	0.4249	0.6516	0.5288	0.5559	0.6035	0.0708	0.0944	0.2266	0.1039	0.1310	0.1786
	VIC	1123	4.2910	3.8246	3.8060	4.0112	3.8308	4.0112	3.9666	0.4851	0.0187	0.2052	0.0249	0.2052	0.1606
9	SA	794	1.7016	1.4398	1.5707	1.6361	1.6667	1.6165	1.7528	0.1309	-0.1309	0.0654	0.0960	0.0458	0.1821
	TAS	518	3.9448	3.6818	4.0631	4.6746	4.1946	4.2702	4.2749	-0.1183	-0.3813	0.6114	0.1315	0.2071	0.2118
	ACT	128	7.4380	6.0606	6.4463	6.6942	7.3278	6.9008	7.0428	0.9917	-0.3857	0.2479	0.8815	0.4545	0.5965
	Aus	14358	2.8326	2.7871	2.8153	2.7530	2.7824	2.7418	2.7440	0.0172	-0.0282	-0.0624	-0.0330	-0.0735	-0.0713
	NSW	4624	3.1006	3.1898	3.0961	3.1363	3.1645	3.1809	3.1549	0.0045	0.0937	0.0402	0.0684	0.0848	0.0588
	QLD	4043	2.7686	2.7178	2.8194	2.7280	2.7686	2.6975	2.6720	-0.0508	-0.1016	-0.0914	-0.0508	-0.1219	-0.1474
	WA	1912	2.4907	2.4377	2.4377	2.4271	2.3671	2.3423	2.3986	0.0530	0.0000	-0.0106	-0.0707	-0.0954	-0.0392
	NT	1292	0.3994	0.5059	0.5751	0.4633	0.4526	0.4673	0.4619	-0.1757	-0.0692	-0.1118	-0.1225	-0.1078	-0.1132
	VIC	1025	4.8806	4.9152	5.0260	4.7560	5.0190	4.8390	4.8241	-0.1454	-0.1108	-0.2700	-0.0069	-0.1869	-0.2018
	SA	820	2.3810	1.7126	2.0551	1.7293	1.7293	1.6291	1.6754	0.3258	-0.3425	-0.3258	-0.3258	-0.4261	-0.3798
	TAS	514	3.9841	3.1208	3.5458	3.3267	3.3333	3.4661	3.4471	0.4382	-0.4250	-0.2191	-0.2125	-0.0797	-0.0987
	ACT	128	7.3171	7.8591	5.6911	6.5041	6.4499	5.2846	6.3803	1.6260	2.1680	0.8130	0.7588	-0.4065	0.6893

Table 43 Estimates of 20th Percentile from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	16700	-1.4698	-1.4574	-1.4612	-1.4675	-1.4660	-1.4660	-1.4633	-0.0086	0.0038	-0.0063	-0.0048	-0.0049	-0.0022
	NSW	5370	-1.2758	-1.2681	-1.2550	-1.2668	-1.2645	-1.2653	-1.2656	-0.0208	-0.0132	-0.0118	-0.0096	-0.0103	-0.0106
	QLD	4723	-1.3142	-1.3256	-1.3097	-1.3320	-1.3249	-1.3256	-1.3197	-0.0045	-0.0160	-0.0223	-0.0153	-0.0159	-0.0100
	WA	2358	-1.7394	-1.7116	-1.7190	-1.7156	-1.7167	-1.7162	-1.7191	-0.0204	0.0074	0.0034	0.0024	0.0029	-0.0001
	NT	1478	-2.2275	-2.2372	-2.2253	-2.2368	-2.2232	-2.2218	-2.2235	-0.0022	-0.0119	-0.0116	0.0021	0.0035	0.0017
	VIC	1164	-1.0147	-1.0108	-1.0080	-1.0302	-1.0307	-1.0249	-1.0213	-0.0067	-0.0028	-0.0222	-0.0227	-0.0169	-0.0133
	SA	899	-1.6570	-1.6872	-1.6425	-1.6437	-1.6335	-1.6535	-1.6392	-0.0145	-0.0447	-0.0012	0.0090	-0.0110	0.0033
	TAS	551	-1.0956	-1.0796	-1.1092	-1.1059	-1.0846	-1.1004	-1.0977	0.0136	0.0297	0.0033	0.0246	0.0088	0.0116
5	ACT	157	-1.0883	-1.1205	-1.1038	-1.1569	-1.1870	-1.1773	-1.1690	0.0156	-0.0167	-0.0531	-0.0832	-0.0735	-0.0652
	Aus	15507	-1.6933	-1.6903	-1.6887	-1.6928	-1.6892	-1.6898	-1.6923	-0.0046	-0.0016	-0.0041	-0.0005	-0.0011	-0.0036
	NSW	4899	-1.4014	-1.4377	-1.4293	-1.4271	-1.4345	-1.4290	-1.4330	0.0280	-0.0084	0.0023	-0.0051	0.0003	-0.0036
	QLD	4508	-1.5541	-1.5496	-1.5406	-1.5482	-1.5408	-1.5475	-1.5487	-0.0135	-0.0090	-0.0076	-0.0002	-0.0069	-0.0081
	WA	2167	-1.9902	-1.9327	-1.9650	-1.9552	-1.9525	-1.9546	-1.9524	-0.0251	0.0323	0.0099	0.0125	0.0104	0.0127
	NT	1440	-2.6226	-2.6832	-2.6778	-2.6984	-2.6978	-2.6925	-2.6833	0.0552	-0.0054	-0.0206	-0.0200	-0.0147	-0.0055
	VIC	1053	-1.2686	-1.1998	-1.2152	-1.2293	-1.2181	-1.2056	-1.2179	-0.0534	0.0154	-0.0141	-0.0029	0.0096	-0.0027
	SA	821	-1.7935	-1.7860	-1.7846	-1.8132	-1.7884	-1.8021	-1.7921	-0.0089	-0.0014	-0.0286	-0.0038	-0.0175	-0.0075
7	TAS	479	-1.2841	-1.2642	-1.2613	-1.2568	-1.2585	-1.2576	-1.2503	-0.0228	-0.0029	0.0044	0.0028	0.0037	0.0110
	ACT	140	-1.4068	-1.2371	-1.3113	-1.2989	-1.3250	-1.3058	-1.3090	-0.0955	0.0743	0.0125	-0.0137	0.0055	0.0024
	Aus	15083	-1.7964	-1.7942	-1.7985	-1.7998	-1.7993	-1.8001	-1.8006	0.0022	0.0043	-0.0012	-0.0008	-0.0016	-0.0020
	NSW	4829	-1.6711	-1.6708	-1.6680	-1.6832	-1.6750	-1.6793	-1.6772	-0.0031	-0.0028	-0.0152	-0.0070	-0.0113	-0.0093
	QLD	4249	-1.6731	-1.6518	-1.6642	-1.6657	-1.6607	-1.6692	-1.6629	-0.0090	0.0124	-0.0016	0.0034	-0.0050	0.0013
	WA	1989	-1.9262	-1.9310	-1.9276	-1.9312	-1.9463	-1.9249	-1.9363	0.0014	-0.0033	-0.0036	-0.0186	0.0027	-0.0087
	NT	1453	-2.7926	-2.8335	-2.7727	-2.8029	-2.7985	-2.8038	-2.7997	-0.0199	-0.0608	-0.0302	-0.0258	-0.0312	-0.0270
	VIC	1123	-1.5501	-1.5465	-1.5526	-1.5386	-1.5438	-1.5446	-1.5446	0.0025	0.0062	0.0141	0.0088	0.0080	0.0080
9	SA	794	-1.7684	-1.7633	-1.7373	-1.7532	-1.7391	-1.7455	-1.7424	-0.0311	-0.0259	-0.0159	-0.0018	-0.0082	-0.0051
	TAS	518	-1.5680	-1.5522	-1.5683	-1.5370	-1.5363	-1.5516	-1.5459	0.0002	0.0161	0.0312	0.0320	0.0166	0.0224
	ACT	128	-1.4258	-1.5033	-1.5788	-1.5481	-1.5415	-1.5816	-1.5816	0.1530	0.0755	0.0307	0.0373	-0.0028	-0.0028
	Aus	14358	-2.0227	-2.0298	-2.0291	-2.0299	-2.0285	-2.0304	-2.0276	0.0064	-0.0008	-0.0009	0.0005	-0.0013	0.0015
	NSW	4624	-1.9306	-1.9535	-1.9411	-1.9440	-1.9483	-1.9449	-1.9447	0.0105	-0.0124	-0.0029	-0.0072	-0.0038	-0.0035
	QLD	4043	-1.9094	-1.9021	-1.9138	-1.9023	-1.9051	-1.9009	-1.9029	0.0043	0.0117	0.0114	0.0086	0.0129	0.0108
	WA	1912	-2.0999	-2.0994	-2.0995	-2.0953	-2.0947	-2.1096	-2.0971	-0.0005	0.0001	0.0041	0.0048	-0.0102	0.0024
	NT	1292	-2.7427	-2.7422	-2.7840	-2.7665	-2.7725	-2.7753	-2.7656	0.0412	0.0418	0.0174	0.0115	0.0087	0.0184
9	VIC	1025	-1.8474	-1.8117	-1.8111	-1.8154	-1.8068	-1.8120	-1.8144	-0.0363	-0.0007	-0.0043	0.0042	-0.0009	-0.0033
	SA	820	-2.1065	-2.0836	-2.0636	-2.0919	-2.0757	-2.0853	-2.0774	-0.0429	-0.0200	-0.0283	-0.0121	-0.0217	-0.0138
	TAS	514	-1.7605	-1.7781	-1.7959	-1.7781	-1.7820	-1.7900	-1.7822	0.0355	0.0178	0.0179	0.0139	0.0059	0.0137
	ACT	128	-1.8771	-1.8690	-1.8283	-1.8483	-1.8581	-1.8551	-1.8870	-0.0487	-0.0407	-0.0200	-0.0298	-0.0268	-0.0587

Table 44 Estimates of 80th Percentile from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs							Difference from 5 PVs					
			1	3	5	10	15	20	46	1	3	10	15	20	46
3	Aus	16700	0.0982	0.0888	0.0958	0.0988	0.0956	0.0962	0.0941	0.0025	-0.0069	0.0030	-0.0002	0.0004	-0.0017
	NSW	5370	0.1964	0.2003	0.2093	0.2160	0.2042	0.2125	0.2078	-0.0129	-0.0090	0.0067	-0.0051	0.0032	-0.0015
	QLD	4723	0.1477	0.1319	0.1313	0.1317	0.1398	0.1352	0.1366	0.0165	0.0006	0.0004	0.0085	0.0039	0.0053
	WA	2358	-0.2120	-0.2282	-0.2371	-0.2153	-0.2373	-0.2300	-0.2322	0.0251	0.0089	0.0218	-0.0002	0.0071	0.0049
	NT	1478	-0.6778	-0.6812	-0.6882	-0.6793	-0.6821	-0.6914	-0.6893	0.0104	0.0069	0.0089	0.0061	-0.0033	-0.0011
	VIC	1164	0.5156	0.4574	0.4710	0.4691	0.4673	0.4692	0.4667	0.0446	-0.0136	-0.0019	-0.0038	-0.0018	-0.0043
	SA	899	-0.2073	-0.1672	-0.1690	-0.1779	-0.1792	-0.1794	-0.1711	-0.0383	0.0017	-0.0089	-0.0102	-0.0104	-0.0021
	TAS	551	0.4360	0.4443	0.4523	0.4683	0.4794	0.4719	0.4598	-0.0162	-0.0080	0.0160	0.0271	0.0196	0.0076
	ACT	157	0.2743	0.3268	0.3186	0.3052	0.2990	0.2788	0.3013	-0.0444	0.0082	-0.0135	-0.0196	-0.0398	-0.0174
5	Aus	15507	0.1016	0.1062	0.1060	0.1119	0.1014	0.1051	0.1054	-0.0044	0.0001	0.0058	-0.0046	-0.0010	-0.0007
	NSW	4899	0.2604	0.2451	0.2550	0.2606	0.2516	0.2581	0.2576	0.0054	-0.0099	0.0056	-0.0034	0.0031	0.0026
	QLD	4508	0.1711	0.1767	0.1697	0.1727	0.1627	0.1644	0.1702	0.0014	0.0070	0.0030	-0.0070	-0.0053	0.0005
	WA	2167	-0.2189	-0.2412	-0.2396	-0.2281	-0.2297	-0.2389	-0.2300	0.0207	-0.0016	0.0115	0.0099	0.0007	0.0096
	NT	1440	-0.8078	-0.8211	-0.8301	-0.8360	-0.8365	-0.8234	-0.8260	0.0223	0.0090	-0.0059	-0.0064	0.0067	0.0041
	VIC	1053	0.4467	0.4452	0.4337	0.4506	0.4456	0.4406	0.4414	0.0130	0.0115	0.0169	0.0119	0.0069	0.0077
	SA	821	-0.2113	-0.2137	-0.1814	-0.1977	-0.1852	-0.1884	-0.1875	-0.0299	-0.0322	-0.0163	-0.0038	-0.0069	-0.0061
	TAS	479	0.5302	0.4727	0.4374	0.5068	0.4414	0.4622	0.4636	0.0928	0.0353	0.0694	0.0040	0.0248	0.0263
	ACT	140	0.2279	0.2655	0.3057	0.3131	0.2998	0.3242	0.3044	-0.0778	-0.0402	0.0074	-0.0059	0.0185	-0.0013
7	Aus	15083	-0.3679	-0.3708	-0.3740	-0.3700	-0.3727	-0.3705	-0.3712	0.0061	0.0032	0.0039	0.0013	0.0035	0.0027
	NSW	4829	-0.2951	-0.3040	-0.2986	-0.2985	-0.3040	-0.2967	-0.2994	0.0035	-0.0054	0.0001	-0.0054	0.0019	-0.0008
	QLD	4249	-0.2869	-0.2650	-0.2696	-0.2699	-0.2641	-0.2659	-0.2663	-0.0174	0.0045	-0.0003	0.0055	0.0037	0.0033
	WA	1989	-0.6254	-0.5811	-0.5993	-0.5858	-0.6013	-0.5933	-0.5893	-0.0260	0.0182	0.0135	-0.0020	0.0060	0.0100
	NT	1453	-1.1700	-1.1579	-1.1618	-1.1504	-1.1577	-1.1587	-1.1569	-0.0082	0.0040	0.0114	0.0041	0.0031	0.0049
	VIC	1123	-0.1614	-0.1422	-0.1356	-0.1543	-0.1437	-0.1326	-0.1418	-0.0258	-0.0066	-0.0187	-0.0081	0.0030	-0.0062
	SA	794	-0.4965	-0.5085	-0.4883	-0.5051	-0.4934	-0.4846	-0.4862	-0.0082	-0.0202	-0.0168	-0.0051	0.0038	0.0021
	TAS	518	-0.2321	-0.2125	-0.2006	-0.1993	-0.1908	-0.2025	-0.1931	-0.0315	-0.0119	0.0014	0.0098	-0.0019	0.0076
	ACT	128	-0.1403	0.0090	-0.1396	-0.0796	-0.0605	-0.0758	-0.0828	-0.0007	0.1486	0.0600	0.0790	0.0638	0.0567
9	Aus	14358	-0.5766	-0.5702	-0.5703	-0.5680	-0.5725	-0.5751	-0.5730	-0.0063	0.0001	0.0023	-0.0021	-0.0048	-0.0027
	NSW	4624	-0.5427	-0.5394	-0.5339	-0.5389	-0.5405	-0.5388	-0.5375	-0.0088	-0.0055	-0.0051	-0.0067	-0.0050	-0.0036
	QLD	4043	-0.5193	-0.5037	-0.5065	-0.5071	-0.5113	-0.5136	-0.5142	-0.0128	0.0028	-0.0006	-0.0048	-0.0071	-0.0077
	WA	1912	-0.6935	-0.6881	-0.6572	-0.6550	-0.6827	-0.6690	-0.6653	-0.0363	-0.0309	0.0022	-0.0255	-0.0118	-0.0081
	NT	1292	-1.2603	-1.2690	-1.2892	-1.2937	-1.2881	-1.2816	-1.2767	0.0289	0.0202	-0.0045	0.0011	0.0076	0.0125
	VIC	1025	-0.3157	-0.2818	-0.2794	-0.2911	-0.2615	-0.2871	-0.2817	-0.0363	-0.0024	-0.0117	0.0179	-0.0077	-0.0023
	SA	820	-0.6545	-0.6836	-0.6620	-0.6626	-0.6821	-0.6953	-0.6877	0.0074	-0.0216	-0.0006	-0.0201	-0.0333	-0.0257
	TAS	514	-0.3592	-0.3963	-0.3940	-0.3724	-0.3793	-0.3957	-0.3845	0.0348	-0.0023	0.0216	0.0146	-0.0017	0.0095
	ACT	128	-0.2575	-0.1570	-0.2022	-0.1412	-0.1404	-0.1361	-0.1527	-0.0554	0.0452	0.0610	0.0617	0.0660	0.0494

Table 45 Standard Error Estimates of Mean from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aust	16700	0.0162	0.0164	0.0168	0.0165	0.0167	0.0166	-0.0002	0.0004	0.0001	0.0003	0.0002
	NSW	5370	0.0182	0.0193	0.0180	0.0185	0.0184	0.0184	-0.0011	-0.0012	-0.0007	-0.0009	-0.0008
	QLD	4723	0.0257	0.0268	0.0268	0.0259	0.0260	0.0261	-0.0011	0.0000	-0.0009	-0.0008	-0.0008
	WA	2358	0.0371	0.0379	0.0389	0.0388	0.0375	0.0378	-0.0008	0.0010	0.0009	-0.0005	-0.0001
	NT	1478	0.0653	0.0662	0.0647	0.0649	0.0658	0.0663	-0.0009	-0.0015	-0.0013	-0.0004	0.0001
	VIC	1164	0.0334	0.0360	0.0354	0.0356	0.0362	0.0351	-0.0026	-0.0006	-0.0004	0.0001	-0.0009
	SA	899	0.0492	0.0545	0.0524	0.0500	0.0484	0.0500	-0.0053	-0.0021	-0.0045	-0.0062	-0.0045
	TAS	551	0.0485	0.0479	0.0543	0.0500	0.0486	0.0499	0.0006	0.0064	0.0021	0.0007	0.0020
	ACT	157	0.0917	0.0907	0.0918	0.0848	0.0847	0.0904	0.0010	0.0011	-0.0060	-0.0061	-0.0003
5	Aust	15507	0.0215	0.0222	0.0220	0.0217	0.0218	0.0219	-0.0007	-0.0002	-0.0005	-0.0004	-0.0003
	NSW	4899	0.0213	0.0223	0.0225	0.0224	0.0221	0.0221	-0.0009	0.0002	0.0001	-0.0002	-0.0002
	QLD	4508	0.0349	0.0346	0.0352	0.0346	0.0353	0.0349	0.0003	0.0006	0.0000	0.0007	0.0003
	WA	2167	0.0446	0.0434	0.0440	0.0436	0.0437	0.0440	0.0012	0.0006	0.0003	0.0003	0.0006
	NT	1440	0.0932	0.0967	0.0953	0.0935	0.0955	0.0949	-0.0035	-0.0014	-0.0033	-0.0012	-0.0018
	VIC	1053	0.0365	0.0397	0.0381	0.0396	0.0397	0.0400	-0.0032	-0.0016	-0.0001	0.0000	0.0003
	SA	821	0.0520	0.0587	0.0581	0.0546	0.0542	0.0542	-0.0067	-0.0007	-0.0041	-0.0045	-0.0046
	TAS	479	0.0572	0.0602	0.0616	0.0601	0.0607	0.0606	-0.0030	0.0014	-0.0001	0.0005	0.0004
	ACT	140	0.1312	0.1022	0.1162	0.0971	0.1065	0.1009	0.0290	0.0140	-0.0051	0.0043	-0.0013
7	Aust	15083	0.0228	0.0224	0.0227	0.0226	0.0226	0.0227	0.0005	0.0003	0.0002	0.0002	0.0003
	NSW	4829	0.0208	0.0207	0.0212	0.0208	0.0213	0.0212	0.0000	0.0005	0.0001	0.0006	0.0005
	QLD	4249	0.0292	0.0291	0.0299	0.0305	0.0299	0.0297	0.0001	0.0008	0.0014	0.0008	0.0006
	WA	1989	0.0356	0.0392	0.0375	0.0358	0.0363	0.0367	-0.0036	-0.0017	-0.0034	-0.0029	-0.0025
	NT	1453	0.1199	0.1176	0.1211	0.1180	0.1184	0.1188	0.0023	0.0035	0.0004	0.0008	0.0012
	VIC	1123	0.0352	0.0414	0.0361	0.0369	0.0370	0.0379	-0.0062	-0.0053	-0.0045	-0.0044	-0.0034
	SA	794	0.0407	0.0428	0.0462	0.0439	0.0439	0.0438	-0.0021	0.0034	0.0011	0.0012	0.0011
	TAS	518	0.0669	0.0498	0.0511	0.0525	0.0522	0.0511	0.0170	0.0013	0.0027	0.0024	0.0013
	ACT	128	0.0855	0.0980	0.1018	0.1037	0.1058	0.0997	-0.0125	0.0038	0.0057	0.0078	0.0017
9	Aust	14358	0.0192	0.0195	0.0187	0.0188	0.0189	0.0188	-0.0003	-0.0008	-0.0006	-0.0006	-0.0007
	NSW	4624	0.0252	0.0226	0.0226	0.0228	0.0230	0.0230	0.0025	0.0000	0.0002	0.0004	0.0004
	QLD	4043	0.0309	0.0310	0.0300	0.0307	0.0307	0.0297	-0.0001	-0.0010	-0.0002	-0.0003	-0.0012
	WA	1912	0.0445	0.0454	0.0443	0.0436	0.0442	0.0433	-0.0009	-0.0011	-0.0018	-0.0012	-0.0021
	NT	1292	0.0850	0.0897	0.0906	0.0868	0.0891	0.0892	-0.0047	0.0009	-0.0029	-0.0007	-0.0006
	VIC	1025	0.0426	0.0462	0.0440	0.0431	0.0471	0.0454	-0.0035	-0.0022	-0.0031	0.0009	-0.0008
	SA	820	0.0511	0.0529	0.0533	0.0520	0.0568	0.0540	-0.0018	0.0004	-0.0009	0.0038	0.0011
	TAS	514	0.0535	0.0655	0.0630	0.0588	0.0571	0.0584	-0.0121	-0.0026	-0.0067	-0.0084	-0.0072
	ACT	128	0.1197	0.1277	0.1334	0.1210	0.1220	0.1257	-0.0081	0.0056	-0.0067	-0.0057	-0.0020

Table 46 Standard Error Estimates of Standard Deviation from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aust	16700	0.0106	0.0102	0.0102	0.0103	0.0101	0.0103	0.0004	0.0001	0.0001	-0.0001	0.0001
	NSW	5370	0.0211	0.0141	0.0103	0.0110	0.0115	0.0116	0.0070	-0.0038	-0.0031	-0.0025	-0.0025
	QLD	4723	0.0115	0.0116	0.0123	0.0122	0.0128	0.0127	-0.0001	0.0007	0.0006	0.0012	0.0011
	WA	2358	0.0241	0.0225	0.0218	0.0228	0.0224	0.0232	0.0016	-0.0006	0.0003	0.0000	0.0007
	NT	1478	0.0383	0.0366	0.0391	0.0370	0.0376	0.0367	0.0017	0.0025	0.0003	0.0009	0.0001
	VIC	1164	0.0261	0.0252	0.0245	0.0253	0.0254	0.0248	0.0009	-0.0007	0.0000	0.0002	-0.0004
	SA	899	0.0353	0.0356	0.0347	0.0325	0.0325	0.0328	-0.0003	-0.0009	-0.0031	-0.0031	-0.0028
	TAS	551	0.0323	0.0342	0.0390	0.0310	0.0319	0.0327	-0.0018	0.0048	-0.0032	-0.0023	-0.0015
	ACT	157	0.0599	0.0748	0.0792	0.0635	0.0644	0.0674	-0.0148	0.0044	-0.0112	-0.0104	-0.0074
5	Aust	15507	0.0148	0.0156	0.0153	0.0149	0.0157	0.0151	-0.0007	-0.0002	-0.0006	0.0001	-0.0005
	NSW	4899	0.0130	0.0136	0.0132	0.0138	0.0149	0.0134	-0.0006	-0.0004	0.0002	0.0013	-0.0003
	QLD	4508	0.0186	0.0197	0.0170	0.0180	0.0166	0.0178	-0.0012	-0.0027	-0.0018	-0.0032	-0.0019
	WA	2167	0.0246	0.0253	0.0269	0.0264	0.0267	0.0260	-0.0007	0.0015	0.0011	0.0013	0.0007
	NT	1440	0.0472	0.0438	0.0456	0.0412	0.0446	0.0436	0.0034	0.0018	-0.0026	0.0008	-0.0002
	VIC	1053	0.0261	0.0397	0.0294	0.0301	0.0277	0.0271	-0.0137	-0.0103	-0.0096	-0.0120	-0.0127
	SA	821	0.0497	0.0432	0.0388	0.0407	0.0422	0.0413	0.0065	-0.0044	-0.0025	-0.0010	-0.0019
	TAS	479	0.0375	0.0427	0.0382	0.0368	0.0377	0.0403	-0.0052	-0.0044	-0.0059	-0.0049	-0.0023
	ACT	140	0.0856	0.0800	0.0772	0.0664	0.0779	0.0753	0.0056	-0.0029	-0.0136	-0.0021	-0.0047
7	Aust	15083	0.0206	0.0197	0.0192	0.0194	0.0193	0.0193	0.0009	-0.0004	-0.0003	-0.0004	-0.0004
	NSW	4829	0.0152	0.0136	0.0129	0.0143	0.0132	0.0137	0.0016	-0.0007	0.0007	-0.0004	0.0001
	QLD	4249	0.0176	0.0152	0.0154	0.0160	0.0161	0.0157	0.0023	0.0002	0.0008	0.0008	0.0005
	WA	1989	0.0175	0.0214	0.0228	0.0176	0.0180	0.0187	-0.0039	0.0015	-0.0038	-0.0034	-0.0027
	NT	1453	0.0505	0.0575	0.0525	0.0523	0.0524	0.0523	-0.0071	-0.0050	-0.0052	-0.0051	-0.0052
	VIC	1123	0.0242	0.0303	0.0247	0.0256	0.0252	0.0261	-0.0061	-0.0056	-0.0047	-0.0051	-0.0042
	SA	794	0.0308	0.0343	0.0326	0.0318	0.0345	0.0337	-0.0035	-0.0017	-0.0025	0.0001	-0.0006
	TAS	518	0.0379	0.0384	0.0358	0.0344	0.0375	0.0358	-0.0005	-0.0026	-0.0040	-0.0009	-0.0026
	ACT	128	0.0722	0.0740	0.0894	0.0821	0.0836	0.0835	-0.0017	0.0154	0.0081	0.0096	0.0095
9	Aust	14358	0.0127	0.0128	0.0133	0.0131	0.0131	0.0130	-0.0001	0.0005	0.0004	0.0004	0.0002
	NSW	4624	0.0200	0.0189	0.0181	0.0182	0.0187	0.0195	0.0011	-0.0008	-0.0007	-0.0002	0.0007
	QLD	4043	0.0187	0.0199	0.0191	0.0178	0.0193	0.0185	-0.0012	-0.0008	-0.0020	-0.0006	-0.0013
	WA	1912	0.0281	0.0271	0.0299	0.0298	0.0301	0.0287	0.0010	0.0029	0.0027	0.0030	0.0016
	NT	1292	0.0387	0.0379	0.0364	0.0379	0.0391	0.0377	0.0007	-0.0015	0.0000	0.0012	-0.0002
	VIC	1025	0.0338	0.0343	0.0336	0.0317	0.0354	0.0338	-0.0005	-0.0007	-0.0026	0.0011	-0.0004
	SA	820	0.0294	0.0360	0.0320	0.0321	0.0301	0.0340	-0.0066	-0.0041	-0.0040	-0.0059	-0.0020
	TAS	514	0.0344	0.0337	0.0365	0.0382	0.0371	0.0368	0.0007	0.0028	0.0045	0.0034	0.0031
	ACT	128	0.0983	0.0961	0.0831	0.0841	0.0919	0.0836	0.0022	-0.0130	-0.0120	-0.0043	-0.0125

Table 47 Standard Error Estimates of Percentage in the Top Band from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aust	16700	0.2364	0.1745	0.1980	0.2249	0.2250	0.2078	0.0619	0.0235	0.0504	0.0505	0.0333
	NSW	5370	0.3900	0.3619	0.3628	0.3719	0.4190	0.3923	0.0281	0.0010	0.0101	0.0571	0.0304
	QLD	4723	0.4381	0.3420	0.3802	0.3986	0.3672	0.3844	0.0961	0.0382	0.0566	0.0252	0.0424
	WA	2358	0.3079	0.3197	0.3329	0.3493	0.3532	0.3505	-0.0117	0.0132	0.0296	0.0336	0.0308
	NT	1478	0.3590	0.2613	0.3106	0.2769	0.3179	0.3108	0.0978	0.0494	0.0156	0.0566	0.0495
	VIC	1164	0.9939	1.1832	1.0173	1.0614	1.0121	1.0413	-0.1893	-0.1659	-0.1218	-0.1711	-0.1419
	SA	899	0.7083	0.7719	0.5679	0.5671	0.5969	0.6106	-0.0636	-0.2040	-0.2049	-0.1750	-0.1613
	TAS	551	1.3627	1.5359	1.7020	1.5344	1.6303	1.5815	-0.1732	0.1662	-0.0015	0.0945	0.0456
	ACT	157	3.8514	2.5841	3.0579	2.3891	2.2349	2.3662	1.2673	0.4738	-0.1950	-0.3492	-0.2179
5	Aust	15507	0.1670	0.1677	0.1831	0.2008	0.1817	0.1903	-0.0007	0.0153	0.0331	0.0140	0.0226
	NSW	4899	0.4493	0.3547	0.3680	0.3704	0.3686	0.3630	0.0946	0.0133	0.0157	0.0139	0.0084
	QLD	4508	0.4005	0.3228	0.3845	0.4082	0.4249	0.3714	0.0777	0.0617	0.0854	0.1021	0.0486
	WA	2167	0.3936	0.3297	0.3914	0.3321	0.3614	0.3334	0.0639	0.0617	0.0024	0.0317	0.0038
	NT	1440	0.2848	0.2099	0.1875	0.1889	0.1974	0.1716	0.0749	-0.0224	-0.0211	-0.0125	-0.0384
	VIC	1053	0.8696	0.8447	0.9180	0.9215	0.8903	0.8719	0.0249	0.0733	0.0768	0.0456	0.0272
	SA	821	0.5608	0.6167	0.6130	0.6471	0.6027	0.5967	-0.0560	-0.0037	0.0303	-0.0141	-0.0201
	TAS	479	1.3805	1.3404	1.2422	1.2372	1.2057	1.3422	0.0401	-0.0981	-0.1032	-0.1347	0.0018
	ACT	140	1.7733	2.1239	2.5050	1.8332	1.9790	1.9706	-0.3506	0.3811	-0.2907	-0.1449	-0.1533
7	Aust	15083	0.1848	0.1708	0.1909	0.1773	0.1809	0.1757	0.0140	0.0201	0.0065	0.0101	0.0049
	NSW	4829	0.3471	0.3316	0.3214	0.3290	0.3225	0.3283	0.0155	-0.0102	-0.0026	-0.0091	-0.0033
	QLD	4249	0.3803	0.3682	0.4269	0.4145	0.3769	0.3939	0.0120	0.0587	0.0462	0.0087	0.0256
	WA	1989	0.2740	0.2873	0.3043	0.3251	0.3290	0.3045	-0.0133	0.0169	0.0378	0.0417	0.0172
	NT	1453	0.2426	0.2516	0.3037	0.2419	0.2629	0.2660	-0.0090	0.0521	-0.0096	0.0113	0.0144
	VIC	1123	0.8905	0.7749	0.6952	0.7197	0.7590	0.7442	0.1156	-0.0798	-0.0552	-0.0160	-0.0307
	SA	794	0.5106	0.5058	0.6341	0.5848	0.6623	0.5738	0.0049	0.1284	0.0790	0.1565	0.0680
	TAS	518	0.9938	0.9955	0.9483	1.0347	1.0648	0.9996	-0.0017	-0.0472	0.0393	0.0693	0.0041
	ACT	128	2.3976	2.6484	2.7138	2.7285	2.5569	2.6396	-0.2509	0.0654	0.0800	-0.0916	-0.0088
9	Aust	14358	0.1917	0.1864	0.1935	0.1925	0.1963	0.1965	0.0053	0.0071	0.0061	0.0099	0.0101
	NSW	4624	0.3509	0.3196	0.3224	0.3170	0.3228	0.3451	0.0313	0.0028	-0.0026	0.0032	0.0255
	QLD	4043	0.3980	0.3901	0.4185	0.3699	0.3854	0.3576	0.0080	0.0284	-0.0202	-0.0046	-0.0325
	WA	1912	0.5172	0.4726	0.5107	0.5118	0.5255	0.4852	0.0445	0.0381	0.0392	0.0528	0.0126
	NT	1292	0.2644	0.2383	0.2852	0.2358	0.2543	0.2404	0.0261	0.0469	-0.0025	0.0160	0.0020
	VIC	1025	1.0393	0.9403	0.8676	0.9541	0.8900	0.8601	0.0989	-0.0728	0.0138	-0.0503	-0.0802
	SA	820	0.5395	0.6076	0.6469	0.5991	0.5762	0.6318	-0.0680	0.0393	-0.0084	-0.0313	0.0243
	TAS	514	0.9995	1.2213	0.9419	1.0318	1.0275	1.0538	-0.2217	-0.2794	-0.1895	-0.1938	-0.1674
	ACT	128	3.4153	3.4831	3.3563	3.1605	3.1405	2.9479	-0.0678	-0.1268	-0.3226	-0.3426	-0.5352

Table 48 Standard Error Estimates of 20th Percentile from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aust	16700	0.1449	0.1475	0.1455	0.1471	0.1449	0.1446	-0.0026	-0.0020	-0.0003	-0.0026	-0.0028
	NSW	5370	0.1491	0.1470	0.1452	0.1483	0.1455	0.1463	0.0021	-0.0018	0.0013	-0.0015	-0.0007
	QLD	4723	0.1684	0.1713	0.1740	0.1769	0.1738	0.1737	-0.0029	0.0027	0.0056	0.0025	0.0024
	WA	2358	0.2326	0.2118	0.2126	0.2117	0.2108	0.2119	0.0207	0.0008	-0.0001	-0.0010	0.0001
	NT	1478	0.2697	0.2645	0.2686	0.2731	0.2718	0.2716	0.0052	0.0041	0.0086	0.0073	0.0071
	VIC	1164	0.1898	0.1938	0.2018	0.2119	0.2062	0.2064	-0.0040	0.0080	0.0181	0.0123	0.0125
	SA	899	0.2404	0.2515	0.2434	0.2445	0.2419	0.2469	-0.0111	-0.0081	-0.0070	-0.0096	-0.0046
	TAS	551	0.2375	0.2299	0.2491	0.2324	0.2438	0.2363	0.0077	0.0193	0.0025	0.0140	0.0064
	ACT	157	0.3481	0.3478	0.3159	0.3250	0.3115	0.3255	0.0003	-0.0318	-0.0227	-0.0362	-0.0223
5	Aust	15507	0.1651	0.1635	0.1662	0.1647	0.1648	0.1646	0.0016	0.0027	0.0012	0.0013	0.0010
	NSW	4899	0.1474	0.1543	0.1546	0.1552	0.1537	0.1551	-0.0070	0.0003	0.0009	-0.0006	0.0008
	QLD	4508	0.2036	0.2026	0.2057	0.1996	0.2010	0.2036	0.0011	0.0031	-0.0030	-0.0016	0.0010
	WA	2167	0.2341	0.2338	0.2403	0.2308	0.2352	0.2326	0.0003	0.0065	-0.0029	0.0014	-0.0012
	NT	1440	0.3387	0.3626	0.3529	0.3423	0.3509	0.3453	-0.0239	-0.0097	-0.0203	-0.0117	-0.0172
	VIC	1053	0.2277	0.2128	0.2127	0.2110	0.2127	0.2114	0.0149	-0.0001	-0.0018	-0.0001	-0.0014
	SA	821	0.2636	0.2460	0.2532	0.2570	0.2524	0.2521	0.0176	0.0073	0.0110	0.0065	0.0062
	TAS	479	0.2756	0.2605	0.2772	0.2716	0.2729	0.2736	0.0151	0.0167	0.0111	0.0124	0.0131
	ACT	140	0.3650	0.3499	0.3836	0.3746	0.3681	0.3607	0.0151	0.0337	0.0247	0.0181	0.0107
7	Aust	15083	0.1559	0.1546	0.1531	0.1546	0.1540	0.1545	0.0013	-0.0016	0.0000	-0.0007	-0.0002
	NSW	4829	0.1444	0.1368	0.1458	0.1420	0.1412	0.1412	0.0077	0.0090	0.0052	0.0044	0.0044
	QLD	4249	0.1723	0.1710	0.1654	0.1650	0.1706	0.1674	0.0013	-0.0056	-0.0060	-0.0004	-0.0036
	WA	1989	0.2033	0.2012	0.1939	0.1903	0.1917	0.1921	0.0021	-0.0073	-0.0109	-0.0094	-0.0091
	NT	1453	0.4559	0.4413	0.4375	0.4368	0.4370	0.4419	0.0146	-0.0038	-0.0044	-0.0043	0.0006
	VIC	1123	0.1870	0.1895	0.1869	0.1916	0.1917	0.1902	-0.0025	-0.0026	0.0021	0.0022	0.0007
	SA	794	0.2097	0.2162	0.2178	0.2224	0.2220	0.2199	-0.0065	0.0016	0.0061	0.0058	0.0037
	TAS	518	0.2102	0.2388	0.2203	0.2232	0.2202	0.2212	-0.0286	-0.0185	-0.0157	-0.0186	-0.0176
	ACT	128	0.3342	0.3353	0.3464	0.3366	0.3231	0.3329	-0.0011	0.0111	0.0013	-0.0122	-0.0024
9	Aust	14358	0.1464	0.1423	0.1438	0.1432	0.1425	0.1435	0.0041	0.0014	0.0008	0.0001	0.0011
	NSW	4624	0.1493	0.1483	0.1471	0.1484	0.1469	0.1496	0.0010	-0.0012	0.0002	-0.0013	0.0013
	QLD	4043	0.1610	0.1726	0.1624	0.1681	0.1651	0.1655	-0.0117	-0.0102	-0.0045	-0.0076	-0.0071
	WA	1912	0.2126	0.2121	0.2113	0.2089	0.2101	0.2133	0.0005	-0.0007	-0.0032	-0.0020	0.0012
	NT	1292	0.3487	0.3426	0.3405	0.3374	0.3406	0.3326	0.0060	-0.0022	-0.0052	-0.0020	-0.0100
	VIC	1025	0.2397	0.2213	0.2235	0.2213	0.2211	0.2208	0.0184	0.0022	0.0001	-0.0002	-0.0004
	SA	820	0.2333	0.2395	0.2256	0.2385	0.2252	0.2273	-0.0062	-0.0139	-0.0010	-0.0143	-0.0122
	TAS	514	0.2235	0.2467	0.2375	0.2464	0.2276	0.2449	-0.0232	-0.0092	-0.0003	-0.0191	-0.0018
	ACT	128	0.3485	0.3589	0.3435	0.3437	0.3447	0.3487	-0.0104	-0.0154	-0.0152	-0.0142	-0.0102

Table 49 Standard Error Estimates of 80th Percentile from Indigenous Students

Year Level	Jurisdiction	Total Number of Students	Number of PVs						Difference from 5 PVs				
			3	5	10	15	20	46	3	10	15	20	46
3	Aust	16700	0.1232	0.1232	0.1242	0.1241	0.1222	0.1237	0.0000	0.0010	0.0009	-0.0010	0.0005
	NSW	5370	0.1497	0.1460	0.1491	0.1462	0.1510	0.1499	0.0037	0.0031	0.0002	0.0050	0.0039
	QLD	4723	0.1683	0.1677	0.1621	0.1675	0.1650	0.1644	0.0006	-0.0057	-0.0003	-0.0027	-0.0033
	WA	2358	0.2027	0.1939	0.1987	0.1966	0.1936	0.1950	0.0088	0.0049	0.0028	-0.0003	0.0011
	NT	1478	0.2814	0.2650	0.2802	0.2735	0.2774	0.2756	0.0164	0.0152	0.0085	0.0124	0.0106
	VIC	1164	0.2191	0.2165	0.2133	0.2116	0.2129	0.2160	0.0026	-0.0032	-0.0049	-0.0036	-0.0005
	SA	899	0.2366	0.2247	0.2254	0.2353	0.2287	0.2314	0.0119	0.0007	0.0106	0.0040	0.0067
	TAS	551	0.2639	0.2624	0.2719	0.2737	0.2723	0.2778	0.0014	0.0095	0.0113	0.0099	0.0153
	ACT	157	0.3575	0.3355	0.3339	0.3479	0.3446	0.3454	0.0220	-0.0016	0.0124	0.0092	0.0099
5	Aust	15507	0.1406	0.1413	0.1415	0.1399	0.1394	0.1390	-0.0007	0.0003	-0.0014	-0.0019	-0.0023
	NSW	4899	0.1706	0.1652	0.1676	0.1642	0.1672	0.1686	0.0054	0.0024	-0.0010	0.0020	0.0035
	QLD	4508	0.1930	0.1940	0.1970	0.2006	0.1979	0.1974	-0.0009	0.0030	0.0066	0.0039	0.0034
	WA	2167	0.2182	0.2208	0.2191	0.2135	0.2160	0.2141	-0.0026	-0.0017	-0.0073	-0.0048	-0.0067
	NT	1440	0.3224	0.3237	0.3244	0.3173	0.3191	0.3169	-0.0014	0.0007	-0.0064	-0.0046	-0.0069
	VIC	1053	0.2299	0.2317	0.2346	0.2304	0.2314	0.2307	-0.0018	0.0030	-0.0013	-0.0002	-0.0010
	SA	821	0.2420	0.2487	0.2511	0.2470	0.2469	0.2431	-0.0068	0.0023	-0.0017	-0.0018	-0.0056
	TAS	479	0.2894	0.2855	0.2858	0.2867	0.2863	0.2891	0.0040	0.0003	0.0013	0.0009	0.0037
	ACT	140	0.3457	0.3764	0.3510	0.3564	0.3598	0.3517	-0.0307	-0.0255	-0.0200	-0.0166	-0.0247
7	Aust	15083	0.1404	0.1391	0.1399	0.1397	0.1405	0.1403	0.0013	0.0008	0.0006	0.0014	0.0012
	NSW	4829	0.1654	0.1721	0.1686	0.1682	0.1699	0.1690	-0.0067	-0.0035	-0.0040	-0.0022	-0.0032
	QLD	4249	0.1899	0.1922	0.1903	0.1895	0.1898	0.1907	-0.0022	-0.0019	-0.0027	-0.0024	-0.0015
	WA	1989	0.2068	0.2062	0.2141	0.2089	0.2132	0.2140	0.0007	0.0079	0.0027	0.0070	0.0078
	NT	1453	0.3220	0.3006	0.3122	0.3091	0.3113	0.3149	0.0214	0.0117	0.0085	0.0107	0.0143
	VIC	1123	0.2264	0.2171	0.2236	0.2233	0.2202	0.2243	0.0092	0.0065	0.0062	0.0031	0.0072
	SA	794	0.2366	0.2410	0.2363	0.2321	0.2390	0.2329	-0.0044	-0.0047	-0.0089	-0.0020	-0.0081
	TAS	518	0.2822	0.2833	0.2722	0.2739	0.2700	0.2729	-0.0011	-0.0110	-0.0094	-0.0133	-0.0104
	ACT	128	0.4680	0.4522	0.4125	0.4278	0.4367	0.4204	0.0158	-0.0398	-0.0244	-0.0155	-0.0318
9	Aust	14358	0.1363	0.1376	0.1365	0.1387	0.1372	0.1371	-0.0013	-0.0011	0.0011	-0.0004	-0.0006
	NSW	4624	0.1617	0.1668	0.1657	0.1636	0.1651	0.1660	-0.0051	-0.0011	-0.0032	-0.0017	-0.0008
	QLD	4043	0.1867	0.1959	0.1862	0.1907	0.1904	0.1900	-0.0092	-0.0097	-0.0053	-0.0056	-0.0060
	WA	1912	0.2337	0.2236	0.2333	0.2316	0.2314	0.2298	0.0101	0.0097	0.0080	0.0078	0.0062
	NT	1292	0.2903	0.2926	0.2992	0.2996	0.2966	0.2981	-0.0023	0.0066	0.0070	0.0040	0.0055
	VIC	1025	0.2333	0.2517	0.2362	0.2444	0.2444	0.2433	-0.0185	-0.0155	-0.0074	-0.0074	-0.0085
	SA	820	0.2676	0.2789	0.2686	0.2714	0.2653	0.2667	-0.0113	-0.0103	-0.0075	-0.0136	-0.0122
	TAS	514	0.2705	0.2855	0.2793	0.2740	0.2826	0.2789	-0.0150	-0.0062	-0.0115	-0.0030	-0.0066
	CT	128	0.4584	0.4690	0.4788	0.4796	0.4759	0.4792	-0.0106	0.0099	0.0106	0.0070	0.0102

Appendix 8 Standard Errors of Mean Estimated by Three Methods

Table 1 Standard Errors of Mean Estimated by Rubin's rule, Analytic Method and Observed Value by Sample Distribution, Test Targeting, Test Length and Sample Size

Distribution	Target	Method	Test Length for Sample Size=200					Test Length for Sample Size=2,000					Test Length for Sample Size=10,000				
			10	20	40	60	80	10	20	40	60	80	10	20	40	60	80
Normal	On-target	Rubin's Rule 3 PVs	0.0873	0.0820	0.0778	0.0757	0.0748	0.0277	0.0260	0.0244	0.0239	0.0235	0.0123	0.0116	0.0109	0.0107	0.0106
Normal	On-target	Analytic Method	0.0935	0.0824	0.0766	0.0747	0.0739	0.0295	0.0261	0.0243	0.0237	0.0234	0.0132	0.0117	0.0109	0.0106	0.0105
Normal	On-target	SD Among Replicates	0.1092	0.0934	0.0839	0.0783	0.0747	0.0338	0.0290	0.0265	0.0245	0.0236	0.0147	0.0128	0.0113	0.0107	0.0103
Normal	Off by 0.5	Rubin's Rule 3 PVs	0.0878	0.0826	0.0781	0.0760	0.0746	0.0275	0.0259	0.0244	0.0238	0.0236	0.0123	0.0115	0.0109	0.0107	0.0105
Normal	Off by 0.5	Analytic Method	0.0937	0.0825	0.0767	0.0748	0.0739	0.0296	0.0261	0.0243	0.0237	0.0234	0.0132	0.0117	0.0109	0.0106	0.0105
Normal	Off by 0.5	SD Among Replicates	0.1127	0.0964	0.0844	0.0797	0.0753	0.0344	0.0302	0.0259	0.0244	0.0248	0.0153	0.0130	0.0117	0.0109	0.0104
Normal	Off by 1.0	Rubin's Rule 3 PVs	0.0879	0.0831	0.0778	0.0759	0.0749	0.0278	0.0258	0.0245	0.0239	0.0236	0.0124	0.0117	0.0110	0.0107	0.0106
Normal	Off by 1.0	Analytic Method	0.0945	0.0830	0.0770	0.0750	0.0740	0.0298	0.0263	0.0244	0.0238	0.0235	0.0133	0.0118	0.0109	0.0106	0.0105
Normal	Off by 1.0	SD Among Replicates	0.1289	0.1030	0.0856	0.0821	0.0771	0.0374	0.0312	0.0267	0.0258	0.0253	0.0163	0.0136	0.0120	0.0109	0.0107
Normal	Off by 2.5	Rubin's Rule 3 PVs						0.0285	0.0268	0.0252	0.0245	0.0241	0.0127	0.0119	0.0113	0.0109	0.0108
Normal	Off by 2.5	Analytic Method						0.0319	0.0277	0.0253	0.0244	0.0239	0.0143	0.0124	0.0113	0.0109	0.0107
Normal	Off by 2.5	SD Among Replicates						0.0549	0.0406	0.0338	0.0303	0.0278	0.0251	0.0174	0.0145	0.0128	0.0122
Uniform	On-target	Rubin's Rule 3 PVs	0.0887	0.0827	0.0783	0.0763	0.0753	0.0276	0.0262	0.0245	0.0240	0.0236	0.0125	0.0116	0.0110	0.0107	0.0106
Uniform	On-target	Analytic Method	0.0928	0.0820	0.0765	0.0746	0.0737	0.0294	0.0260	0.0242	0.0236	0.0234	0.0131	0.0116	0.0108	0.0106	0.0105
Uniform	On-target	SD Among Replicates	0.1130	0.0948	0.0862	0.0835	0.0802	0.0369	0.0311	0.0292	0.0268	0.0246	0.0160	0.0138	0.0128	0.0124	0.0113
Uniform	Off by 0.5	Rubin's Rule 3 PVs	0.0885	0.0831	0.0781	0.0764	0.0751	0.0280	0.0263	0.0246	0.0240	0.0237	0.0124	0.0117	0.0110	0.0107	0.0106
Uniform	Off by 0.5	Analytic Method	0.0930	0.0822	0.0766	0.0746	0.0737	0.0294	0.0260	0.0243	0.0237	0.0234	0.0132	0.0116	0.0109	0.0106	0.0105
Uniform	Off by 0.5	SD Among Replicates	0.1141	0.0973	0.0885	0.0855	0.0808	0.0342	0.0295	0.0252	0.0248	0.0227	0.0159	0.0143	0.0131	0.0118	0.0115
Uniform	Off by 1.0	Rubin's Rule 3 PVs	0.0890	0.0837	0.0785	0.0769	0.0754	0.0284	0.0263	0.0248	0.0241	0.0238	0.0127	0.0117	0.0111	0.0108	0.0106
Uniform	Off by 1.0	Analytic Method	0.0936	0.0826	0.0768	0.0748	0.0739	0.0296	0.0262	0.0243	0.0237	0.0234	0.0132	0.0117	0.0109	0.0106	0.0105
Uniform	Off by 1.0	SD Among Replicates	0.1227	0.1002	0.0922	0.0868	0.0822	0.0383	0.0319	0.0284	0.0276	0.0275	0.0166	0.0142	0.0132	0.0124	0.0117
Uniform	Off by 2.5	Rubin's Rule 3 PVs						0.0288	0.0272	0.0257	0.0251	0.0245	0.0129	0.0122	0.0115	0.0112	0.0110
Uniform	Off by 2.5	Analytic Method						0.0317	0.0275	0.0251	0.0243	0.0239	0.0142	0.0123	0.0113	0.0109	0.0107
Uniform	Off by 2.5	SD Among Replicates						0.0563	0.0426	0.0351	0.0323	0.0310	0.0245	0.0190	0.0152	0.0141	0.0138
Bimodal	On-target	Rubin's Rule 3 PVs	0.0894	0.0831	0.0787	0.0764	0.0756	0.0280	0.0261	0.0246	0.0240	0.0236	0.0125	0.0116	0.0110	0.0107	0.0106
Bimodal	On-target	Analytic Method	0.0927	0.0820	0.0765	0.0747	0.0738	0.0294	0.0260	0.0242	0.0237	0.0234	0.0131	0.0116	0.0108	0.0106	0.0105
Bimodal	On-target	SD Among Replicates	0.1101	0.0915	0.0847	0.0806	0.0798	0.0352	0.0305	0.0258	0.0265	0.0241	0.0150	0.0128	0.0119	0.0108	0.0106
Bimodal	Off by 0.5	Rubin's Rule 3 PVs	0.0893	0.0839	0.0784	0.0764	0.0756	0.0279	0.0260	0.0246	0.0241	0.0237	0.0125	0.0117	0.0110	0.0107	0.0106
Bimodal	Off by 0.5	Analytic Method	0.0928	0.0821	0.0765	0.0747	0.0738	0.0294	0.0260	0.0243	0.0237	0.0234	0.0132	0.0116	0.0109	0.0106	0.0105
Bimodal	Off by 0.5	SD Among Replicates	0.1158	0.0936	0.0859	0.0818	0.0823	0.0352	0.0306	0.0262	0.0251	0.0245	0.0157	0.0129	0.0121	0.0110	0.0107
Bimodal	Off by 1.0	Rubin's Rule 3 PVs	0.0905	0.0841	0.0790	0.0767	0.0758	0.0281	0.0263	0.0248	0.0241	0.0238	0.0127	0.0117	0.0111	0.0108	0.0106
Bimodal	Off by 1.0	Analytic Method	0.0934	0.0825	0.0768	0.0749	0.0739	0.0296	0.0262	0.0243	0.0237	0.0235	0.0132	0.0117	0.0109	0.0106	0.0105
Bimodal	Off by 1.0	SD Among Replicates	0.1264	0.1013	0.0886	0.0829	0.0812	0.0366	0.0312	0.0275	0.0258	0.0250	0.0160	0.0132	0.0122	0.0109	0.0109
Bimodal	Off by 2.5	Rubin's Rule 3 PVs						0.0290	0.0273	0.0256	0.0250	0.0246	0.0128	0.0122	0.0115	0.0111	0.0110
Bimodal	Off by 2.5	Analytic Method						0.0318	0.0275	0.0252	0.0243	0.0239	0.0142	0.0123	0.0113	0.0109	0.0107
Bimodal	Off by 2.5	SD Among Replicates						0.0560	0.0414	0.0332	0.0302	0.0282	0.0243	0.0179	0.0152	0.0132	0.0126

Appendix 8

Standard Errors of Mean Estimated by Three Methods

Chisq(5)	On-target	Rubin's Rule 3 PVs	0.0853	0.0801	0.0760	0.0738	0.0733	0.0270	0.0249	0.0239	0.0233	0.0230	0.0121	0.0112	0.0107	0.0104	0.0103
Chisq(5)	On-target	Analytic Method	0.0947	0.0831	0.0771	0.0752	0.0743	0.0299	0.0263	0.0244	0.0238	0.0236	0.0134	0.0118	0.0109	0.0107	0.0105
Chisq(5)	On-target	SD Among Replicates	0.1120	0.0927	0.0859	0.0816	0.0815	0.0353	0.0292	0.0267	0.0239	0.0238	0.0156	0.0130	0.0120	0.0113	0.0116
Chisq(5)	Off by 0.5	Rubin's Rule 3 PVs	0.0874	0.0814	0.0765	0.0747	0.0737	0.0274	0.0255	0.0241	0.0235	0.0231	0.0124	0.0113	0.0108	0.0105	0.0104
Chisq(5)	Off by 0.5	Analytic Method	0.0940	0.0829	0.0771	0.0753	0.0744	0.0297	0.0262	0.0244	0.0239	0.0236	0.0133	0.0117	0.0109	0.0107	0.0106
Chisq(5)	Off by 0.5	SD Among Replicates	0.1122	0.0913	0.0861	0.0811	0.0816	0.0342	0.0289	0.0259	0.0262	0.0237	0.0158	0.0132	0.0118	0.0116	0.0114
Chisq(5)	Off by 1.0	Rubin's Rule 3 PVs	0.0886	0.0830	0.0779	0.0756	0.0742	0.0280	0.0261	0.0246	0.0238	0.0234	0.0125	0.0116	0.0110	0.0106	0.0104
Chisq(5)	Off by 1.0	Analytic Method	0.0940	0.0830	0.0774	0.0755	0.0746	0.0297	0.0263	0.0245	0.0239	0.0237	0.0133	0.0118	0.0110	0.0107	0.0106
Chisq(5)	Off by 1.0	SD Among Replicates	0.1221	0.0991	0.0873	0.0853	0.0838	0.0355	0.0304	0.0272	0.0250	0.0244	0.0165	0.0138	0.0121	0.0114	0.0116
Chisq(5)	Off by 2.5	Rubin's Rule 3 PVs						0.0315	0.0289	0.0265	0.0254	0.0246	0.0139	0.0128	0.0118	0.0112	0.0110
Chisq(5)	Off by 2.5	Analytic Method						0.0306	0.0272	0.0252	0.0245	0.0241	0.0137	0.0122	0.0113	0.0110	0.0108
Chisq(5)	Off by 2.5	SD Among Replicates						0.0521	0.0407	0.0314	0.0289	0.0285	0.0231	0.0174	0.0154	0.0136	0.0125