Dialect Speech and Wages*

Yuxin Yao[†] and Jan C. van Ours[‡]

January 17, 2019

Abstract

Our paper studies the causal effect of dialect speech on wages of native Dutch workers. Using an instrumental variable approach, we find evidence of a wage penalty of dialect speech for males but not for females.

Keywords: Dialect speech, Wage penalty, Job characteristics

JEL code: J24, I2

*We thank CentER Data for making the LISS data available. We declare no conflict of interest.

[†]Department of Economics, Faculty of Economics and Management, East China Normal University, China; yxyao@fem.ecnu.edu.cn

[‡]Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; Tinbergen Institute (Amsterdam/Rotterdam); Department of Economics, University of Melbourne, Parkville, Australia; CEPR (London), IZA (Bonn); vanours@ese.eur.nl.

1 Introduction

Labor market outcomes for individual workers may be affected by their language skills. Previous studies have focused on how language skills in host countries affect the labor market position of immigrants. Recent examples are Bleakley and Chin (2010), Miranda and Zhu (2013) and Yao and van Ours (2015). Very few studies focus on how speech patterns affect labor market outcomes of native residents. Grogger (2011, 2019) analyzed audio data from NLSY finding that wages are strongly associated with speech patterns among both African Americans and Southern whites. In this paper, we study the relationship between a regional dialect speech pattern and labor market outcomes of native Dutch residents.

The Netherlands is a small country in which there is limited spatial segregation and the population is homogeneous in terms of culture, physical characteristics and economic wealth. The differences between a language and a dialect are not straightforward. A language often consists of a variety of slightly different dialects in terms of vocabulary, grammar and pronunciation of which one is considered to be the standard language. In the Netherlands, Standard Dutch is spoken in its purest form in Haarlem, the capital municipality of the province of Noord Holland. But the formal status of regional variations of Standard Dutch differs from province to province. Frisian, mostly spoken in the province of Friesland, is recognized as a co-official language with Dutch. Limburgish, spoken in the province of Limburg has a semi-official status, while variations of Dutch spoken in other regions of the Netherlands have a dialect status (see an overview in Driessen (2005)). Dialects are usually acquired without formal education while Standard Dutch is the instruction medium at schools. Speaking with a local dialect accent may signal lower language ability, limited educational skills and thus low productivity. Although in some areas speaking a dialect can be viewed as a separate skill, the return is somewhat limited in other areas of the country. Yao and van Ours (2019) find a negative conditional correlation between hourly wages and daily dialect-speaking behavior. But speech patterns can be endogenous to labor market outcomes. In this paper, we use the geographic distance from the residing municipality to Haarlem as an instrumental variable to individual dialect speech pattern. When we control

for personal, family and province characteristics and account for potential endogeneity, we find a significant causal wage penalty of dialect speech for males, but not for females.

2 Data

The dataset is from the Longitudinal Internet Studies for the Social sciences (LISS), a panel survey representative of the Dutch population (see for details: www.lissdata.nl). We use seven waves of panel data from 2008 to 2014 on Dutch natives aged 15 to 64. Similar to most literature on language effects, we rely on self-reported information. Respondents indicate their dialect speech pattern by answering the question *Do you ever speak dialects?* with possible answers *Yes – daily*, *Yes – regularly*, *Yes – once a while* and *No – never*. The indicator for dialect speech we use in our analysis is whether or not an individual speaks a dialect irrespective of the intensity, in the sense that it is more likely to be predetermined by the regional origin. We refer also to Frisian as a dialect although we are aware of its official status. We use an alternative survey question *Which language do you generally speak at home?* for respondents from Friesland.

We investigate whether dialect speech has a causal effect on wages of Dutch workers. Our dialect speech indicator is limited in the sense that it is not clear which dialect is spoken or whether one speaks a dialect at the workplace or at home. Also, we have no information about the perception of others on whether the speech pattern of the worker sounds like Standard Dutch or whether the worker has a dialect accent such that the regional origin of the worker is revealed.

As explained in more detail in the next section, we use the geographic distance to Haarlem as an instrumental variable in search for causal effects. Figure 1 shows the positive association at the provincial level between the average distance to Haarlem and the share of dialect speakers in our sample.

– Figure 1 about here –

3 Set-up of the Analysis

From our data, we find that over our period of analysis, only 5-6% of individuals ever changed dialect-speaking behavior. Therefore, it is not feasible to use an individual fixed-effects approach to account for unobserved heterogeneity. Instead, we use a pooled cross-section set-up relating dialect speech to labor market outcomes:

$$\log(w_{it}) = \alpha_t + \gamma D_{it} + \beta_1 x_{it} + u_{it} \quad (1)$$

where w_{it} represents hourly wages of individual i in year t , the α_t indicate calendar year fixed effects, D_{it} is a dummy variable for dialect speech, x_{it} is a vector of individual, family and province characteristics, γ is the main parameter of interest, β_1 represents a vector of parameters and u_{it} is an error term. The calendar year fixed effects account for general calendar-time related developments in wages including wage inflation. The personal characteristics include age, educational attainment, having a religion and having a partner. Family characteristics are number of children and the urbanization level of the residence, and province characteristics are log of per capita GDP, log of population, area of main roads in km² and log of employment.

When studying the relationship between dialect speech and wages, there are several threats to identification of a causal effect. First, job characteristics can reversely determine dialect use. Workers in low-skilled occupations for example may not be required to speak Standard Dutch. Second, unobserved factors may cause spurious correlation, for example when low-ability workers combine low productivity with dialect speech. Third, there may be survey measurement errors in establishing the definition of dialects and whether or not individuals speak a dialect. To investigate causal effects of dialect speech on wages, we employ an instrumental variable approach. We assume that dialect speech depends on observed characteristics and the geographical distance of the residing municipality to Haarlem, z_{it} .

$$D_{it} = \alpha_{2t} + \theta z_{it} + \beta_2 x_{it} + \nu_{it} \quad (2)$$

where α_{2t} are year fixed effects and ν_{it} is an error term. Among the observed characteristics are economic factors measured at a provincial level. Our identifying assumption is that conditional on all factors, the distance to Haarlem does not directly affect labor market outcomes of workers but only indirectly through dialect speech.

Our identifying assumption also implies that people do not migrate for work to provinces with a different dialect. A back-of-the-envelope calculation shows that this assumption is not too strong at least for younger cohorts. According to Statistics Netherlands, the annual migration rate between provinces is as low as 1.4-1.8% for the last two decades. Suppose children never move without parents, the province of current residence is very likely to be the province of origin. For instance, for a 35-year-old in 2014, the probability of staying in the province of residence at 15 years old is 72.6%. Moreover, this migration rate includes frequent moves of the same individual, i.e., many individuals move between provinces several times throughout their life or even migrate back to their province of origin. For an individual the probability of staying in the province of origin throughout life is even higher. Finally, by way of sensitivity analysis, we regressed the net population inflow at the province level on economic factors finding that they are uncorrelated after province fixed effects and year fixed effects are taken into account (see Web-appendix 1 for details). Therefore, the province of current residence is pretty much exogenous as it is predetermined by regional origin.

– Table 1 about here –

4 Parameter Estimates

We first relate dialect speech to the geographical distance between the residing municipality and Haarlem. As shown in Table 1 when personal characteristics, family characteristics and provincial characteristics are controlled for, the geographic distance to Haarlem has a significant positive effect on the probability of dialect speech. For every 10 kilometers closer to Haarlem, the probability of dialect-speaking increases with 4.4% for males and 3.7% for females. The F-statistics for excluding the geographical distance are very high, indicating

that if we use this in an instrumental variable approach, our estimates do not suffer from a weak instrument.

– Table 2 about here –

Table 2 shows the relevant parameter estimates of the effect of dialect speech on log hourly wages. Row a shows the results from an OLS estimate indicating significant lower wages for dialect speaking males of 3.2%. For females, the negative effect is 2.7% although this estimate is not significantly different from zero.

Panel b of Table 2 reports the 2SLS estimates for hourly wages where we use the geographical distance to Haarlem as an instrumental variable to account for possible endogeneity of dialect speech. As shown in panel b1, dialect speech significantly decreases males' hourly wages by 11.9%. Dialect speech does not have much impact on females' hourly wages, and the effect does not differ significantly from zero. Conditional on all personal, family and province characteristics the 2SLS estimates suggest a causal significant negative effect of dialect speech on wages of males, consistent with the OLS estimates. Previous studies on language effects using an IV strategy (Bleakley and Chin, 2004; Dustmann and Fabbri, 2003; Dustmann and van Soest, 2001), usually find that 2SLS parameter estimates are larger than OLS parameter estimates. It indicates that the potential upward bias of unobserved heterogeneity and reverse causality is dominated by the downward bias from measurement errors. Our results are in line with this, i.e. the wage penalty effect is underestimated by the OLS estimation which ignores the measure error of self-reported speech pattern. For instance, respondents who report not to speak a dialect could be perceived by employers to have a regional accent.

Panels b2 to b6 show the relevant parameter estimates for a range of sensitivity checks. In panel b2, we exclude individuals who self-report having problems in speaking Standard Dutch. Both for males and females, the parameter estimates are almost the same as in panel b1. This suggests that worse language skills in the standard language are not a reason for the wage penalty of dialect speech. In panel b3, we exclude individuals from Friesland where the regional language has an official status and may therefore bring some benefits. Again,

the size of the wage penalty is very similar to our baseline estimates. Our main conclusion is not confounded by our definition of dialect. Panel b4 investigates whether the effect of dialect speech exists among the cohorts younger than 35 years old who are more likely to have stayed in the province of origin. We add the interaction of dialect speech and a dummy for the younger cohorts, and instrument it by the interaction of geographic distance to Haarlem and this cohort dummy. Dialect speech has no additional effect on the younger cohorts. The wage penalty exists for males in both the young and the old cohorts, suggesting that our conclusion is not biased by internal migration.

Panels b5 and b6 shed some light upon potential mechanisms of the wage penalty of dialect speech. In panel b5, we add the interaction of dialect speech and a dummy for obtaining a college or higher degree, and instrument it by the interaction of geographic distance to Haarlem and the degree dummy. We find that for both genders there is a larger wage penalty of dialect speech for the high-educated, while dialect speech is less a concern for the low-educated. This may be because the high-educated are working in jobs that require communication and travel across the country. It may also be that high-educated dialect speakers are perceived by employers to have a lower unobserved ability. In panel b6, we investigate how dialect speech affects the job type of the workers. We find that speaking a dialect significantly reduces the probability of having a non-manual job for males, but the penalty is not significant and smaller in magnitude for females. Thus the wage penalty of dialect speech for males works at least to some extent through occupational sorting.

5 Conclusions

Using data from the Netherlands, our analysis shows that male workers who speak a dialect have significantly lower hourly wages. High-educated male workers are more heavily penalized. Our data do not allow us to make a distinction between various mechanisms. We speculate that speech-based discrimination from employers or unobserved ability signaled by dialect speech may be a reason. The wage penalty may be partly related to occupational sorting when dialect-speaking workers self-select into jobs located in the province of origin

which require less commuting and communication. This may also explain the gender heterogeneity of the wage penalty of dialect speech, i.e., males are penalized by dialect speech but females are not. Based on our sample females are likely to choose a closer workplace to home than males. Distance from work to home and daily commuting time is 10.5 km and 20.6 minutes on average for females, compared with 16.3 km and 27.5 minutes for males. Therefore, there should be less wage penalty and even some benefits of dialect speech for females who do not commute for work to other provinces where a different dialect or the standard language is spoken.

References

- Bleakley, H. and A. Chin (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics* 86(2), 481–496.
- Bleakley, H. and A. Chin (2010). Age at arrival, english proficiency, and social assimilation among U.S. immigrants. *American Economic Review: Applied Economics* 2(1), 165–192.
- Driessen, G. (2005). In Dutch? Usage of Dutch regional languages and dialects. *Language, Culture and Curriculum* 18, 271–285.
- Dustmann, C. and F. Fabbri (2003). Language proficiency and labour market performance of immigrants in the UK. *Economic Journal* 113(489), 695–717.
- Dustmann, C. and A. van Soest (2001). Language fluency and earnings: Estimations with misspecified indicators. *Review of Economics and Statistics* 83(4), 663–674.
- Grogger, J. (2011). Speech patterns and racial wage inequality. *Journal of Human Resources* 46, 1–25.
- Grogger, J. (2019). Speech and wages. *Journal of Human Resources*. forthcoming.
- Miranda, A. and Y. Zhu (2013). English deficiency and the native immigrant wage gap. *Economics Letters* 118(1), 38–41.
- Yao, Y. and J. C. van Ours (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics* 34, 76–85.
- Yao, Y. and J. C. van Ours (2019). Daily dialect-speaking and wages among native Dutch speakers. *Empirica*, forthcoming.

Figure 1: Relationship between dialect speech and geographic distance to Haarlem

Source: Longitudinal Internet Studies for the Social sciences

Table 1: Parameter estimates of dialect speech

	Males		Females	
Distance to Haarlem/10	0.044	(0.003)***	0.037	(0.003)***
F-statistic	232.6		116.1	

Note: The estimation is based on 6311 observations of 1858 males and 6363 observations of 1897 females. Personal characteristics, family characteristics, province characteristics and year fixed effects are included in all regressions. In parentheses, standard errors clustered at the individual level. Full estimation results are presented in Web-appendix 2.

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Table 2: Parameter estimates of log hourly wages; effect of dialect speech

	Males		N	Females		N
a. OLS	-0.032	(0.016)**	6311	-0.027	(0.017)	6363
b. 2SLS						
1. Baseline	-0.119	(0.055)**	6311	-0.004	(0.068)	6363
2. Exclude language problems	-0.111	(0.057)*	5720	0.001	(0.074)	5594
3. Exclude Friesland	-0.116	(0.053)**	6027	-0.007	(0.067)	6064
4. Dialect speech	-0.117	(0.058)**		0.004	(0.070)	
Dialect speech $\times$ Young cohorts	-0.013	(0.064)	6311	0.002	(0.069)	6363
5. Dialect speech	-0.048	(0.057)		0.039	(0.069)	
Dialect speech $\times$ High-educated	-0.177	(0.071)**	6311	-0.095	(0.075)	6363
6. Non-manual job	-0.158	(0.067)**	6311	-0.074	(0.065)	6363

Note: N = number of observations. Personal characteristics, family characteristics, province characteristics and year fixed effects are controlled in all regressions. In the 2SLS estimates, the geographical distance to Haarlem is used as an instrumental variable. In parentheses, standard errors clustered at the individual level. Full estimation results of rows a and b1 are presented in Web-appendix 2.

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.