The Wage Penalty of Dialect-Speaking*

Yuxin Yao† and Jan C. van Ours‡

January 31, 2016

Abstract

This paper studies the effects of dialect-speaking on job characteristics of Dutch workers, in particular on their wages. Using the geographic distance to Amsterdam as an instrumental variable, we find that male workers who speak a dialect daily on average have significantly lower monthly earnings and hourly wage. For females we also find wage penalty of dialect-speaking but not significantly different from zero.

Keywords: Dialect, Wage penalty, Job characteristics

JEL code: J24, I2

*We are grateful to CentERData for making the LISS data available for this paper.

†Department of Economics, CentER, Tilburg University, The Netherlands; y.yao@uvt.nl

‡Department of Economics, CentER, Tilburg University, The Netherlands; Department of Economics, University of Melbourne, Parkville, Australia; CEPR(London), CESifo(Munich), CREAM (London), IZA(Bonn); vanours@uvt.nl.

1 Introduction

Language skills are an important determinant of labor market performance. Previous studies have focused on the effect of language proficiency on earnings of male immigrants. Recent examples are Miranda and Zhu (2013a), Miranda and Zhu (2013b), Budría and Swedberg (2012), Di Paolo and Raymond (2012) and Yao and van Ours (2015). However, it is not only language proficiency that affects labor market performance. Also, language speech patterns may be important, i.e. it may matter whether a worker speaks a standard language or a dialect. Even among linguists there is no common definition of dialect as distinct from a standard language. There is a view that a standard language is the language of the majority of people but there is also a view that the standard language is for the selected few (Smakman (2012)). A dialect may refer to a regional speech pattern. However, a dialect may also refer to social class as for example is apparent from the "My Fair Lady" lyrics: "Look at her, a prisoner of the gutter, condemned by every syllable she utters (...). An Englishman's way of speaking absolutely classifies him. The moment he talks he makes some other Englishman despise him."[1]

To study the effects of speech patterns, Grogger (2011) uses NLSY data in combination with audio-information about how individuals speak. It turns out that speech pattern affect wages. In the US labor market, black workers with a distinct black speech earn less than white workers whereas black workers who do not sound distinct black earn the same as white workers. Gao and Smyth (2011) find a significant wage premium associated with fluency in standard Mandarin for dialect-speaking migrating workers in China. Carlson and McHenry (2006) presents the results of a small experiment on how speaking dialect affects employment probability. Bendick Jr. et al. (2010) using an experimental set-up studied the effects of a (mostly French accent) for white job applicants to New York City restaurants. These accents were considered as "charming" and they increased the probability of being hired as a waiter or waitress. According to Das (2013) language and accents provide information about an individual's social status. The spoken language may be a source of discrimination affecting earnings and promotion.

Grogger (2011) suggests two possible explanations for the effect of speech on earnings. First, there could be a causal effect working either through productivity or through discrimination. Language speech differences among workers may increase production costs (Lang

[1] From the song "Why can't the English?"

(1986)) or language speech may influence the severity of discrimination by employers. A second explanation is that speech is a signal of unobserved productivity.

Our paper studies the relationship between dialect-speaking and labor market performance of Dutch workers, in particular in the wage effects.[2] There is a wide variation in the dialects spoken in the Netherlands. Standard Dutch originated in the urban areas of Noord-Holland, Zuid-Holland and Utrecht. Frisian is recognized as a separate language and promoted by the local government. In the province Friesland, Frisian is spoken by a majority of the population. Almost all inhabitants understand the language and can read it but only a small minority can write the language. All Frisian-speakers are able to speak, read, and write Dutch. The Frisian language is officially recognized as the second language of the Netherlands (Gorter (2005)). Although officially Frisian is a language we refer to it as a dialect. There are also the regional languages Limburgish and Low Saxon. Limburgish is spoken in the province of Limburg by about 75% inhabitants and Low Saxon is spoken in the provinces of Groningen, Drenthe, Overijssel and Gelderland by approximately 60% inhabitants. Regional languages have an official status in the related regions but they are no clear regulations regarding to the support of government. Furthermore, there are a few Dutch dialects, such as Brabantish spoken in Noord-Brabant and Zeelandic in Zeeland.

Our paper is set-up as follows. In section 2 we provide a description of our data. Section 3 presents our statistical model and discusses our identification strategy. Section 4 discusses our parameter estimates. Section 5 concludes.

2 Our Data

Our dataset is from the LISS (Longitudinal Internet Studies for the Social sciences) survey. In this survey, background demographic variables are collected monthly while on specific topics data are collected annually. We use the available 7 waves of panel data from 2008 to 2014 and focus on four indicators of labor market performance: employment, working hours, earnings and type of jobs. An individual is considered to be employed if he or she has any type of paid work, including family business and self-employment. And based on average weekly working hours and personal monthly gross earnings, we can calculate hourly wages.

[2] Yao et al. (2016) investigate the relationship between dialect speaking and academic performance of 5-6 year old children in the Netherlands. They find that dialect speaking has a modestly negative effect on language skill with larger effects on boys.

As is usual in studies on language effects, we rely on self-reported information. Respondents indicate dialect-speaking by answering the questions: *Do you ever speak dialect?* Respondents can choose between *Yes, daily*, *Yes, regularly*, *Yes, once a while* and *No, never*. The indicator for speaking dialect we use in our analysis is defined as a dummy variable which equals 1 if the individual speaks a dialect daily[3] To focus on the effects of dialect speaking pattern and exclude the effect of language deficiency, we remove from the sample immigrants and individuals who indicate having problems in reading or speaking Dutch. Our sample consists of 22,581 observations from 6,907 respondents. We merged the dataset with a variable measuring the geographic distance from residential municipality to Amsterdam from an additional confidential dataset. Table 1 provides information on linguistic background by provinces. As shown, the linguistic distance to standard Dutch of the dialect spoken in a particular province is the largest in Friesland and the smallest in Flevoland, Noord-Holland and Zuid-Holland. The distance to Amsterdam (measured as multiples of 5 kilometers) is on average the largest in Gelderland and Groningen and the smallest in Noord-Holland in which Amsterdam is located. Dialect-speaking is the most prevalent in Limburg where 68% of the individuals in our sample that speak dialect daily, followed by Friesland with 48% and Drenthe with 34%. Flevoland, Noord-Holland, Utrecht and Zuid-Holland only have a tiny proportion of sample speaking dialect daily. Plus, Noord-Holland and Zuid-Holland are the provinces with the highest share of individuals in our sample who never speak dialect.

Table 2 shows the summary statistics by gender and daily dialect-speaking. Comparing dialect speakers with Dutch speakers, we will find that dialect speakers have slightly lower education and more difficulty in Dutch, and are much less likely to live in urbanized areas. We do not find much difference in province characteristics between the two groups, except that higher linguistic distance score of local language to standard Dutch. If we look at labor market characteristics, employment rate and weekly working hours are similar between dialect speakers and Dutch speakers. However, dialect speakers have a lower hourly wage and lower monthly earnings.

We plot the densities and cumulative distributions of monthly wage, hourly wage and working hours by the dialect status in Figures 1 and 2. The top graph of Figure 1 shows that

[3]Since Frisian is usually viewed as a separate language rather than a dialect, we refer to the survey question "*which language do you generally speak at home?*" for respondents from Friesland. In this paper, we do not distinguish Frisian with Dutch dialects. Frisian speakers at home are considered to speak dialects daily.

Table 1: Linguistic characteristics by province

Province	Linguistic distance	Distance to Amsterdam	Speaking dialect (%) Daily	Regularly	Sometimes	Never	N
Drenthe	19	25.7	33.9	11.6	17.1	37.4	655
Flevoland	12	7.6	1.0	5.3	10.7	83.0	495
Friesland	37	21.2	47.6	9.5	13.4	29.6	1,110
Gelderland	28	16.4	14.3	9.2	19.5	57.1	3,106
Groningen	28	32.1	21.5	12.5	20.5	45.6	904
Limburg	32	32.0	67.8	7.1	9.6	15.5	1,628
Noord-Brabant	28	19.1	22.4	15.0	27.0	35.6	3,983
Noord-Holland	12	4.1	2.8	1.8	8.6	86.9	2,940
Overijssel	29	22.2	25.1	15.5	27.6	31.8	1,545
Utrecht	18	7.7	3.0	4.1	11.0	81.9	1,594
Zeeland	29	25.7	29.2	15.5	26.0	29.2	489
Zuid-Holland	12	10.3	3.3	2.2	9.4	85.2	4,127
Total	22.5	16.2	18.5	8.2	16.4	56.9	22,576

Source: Linguistic distance: Van Bezooijen and Heeringa (2006)

there are clear differences between individuals who speak dialect daily and those that do not.[4] Monthly earnings are lower for females irrespective of whether or not they speak dialect. For both males and females there is more mass at the lower end of the wage distribution for individuals who speak dialect daily. The differences in monthly earnings between males and females is partly related to differences in weekly working hours. As shown in the middle graph of Figure 1 the peak in the weekly working hours for males is around 40. For females, the distribution is spread-out more. There is a peak around 20 but there are also peaks around 30 and 40. For males there is hardly any difference in the distribution of weekly working hours between those who speak dialect daily and those who do not. For women there are some differences with women who speak dialect daily working fewer hours. The bottom graph of Figure 1 shows the distribution of hourly wages. The differences between males and females are not big but also here individuals who speak dialect daily on average have lower hourly wages. The graphs in Figure 2 show that workers who speak dialect daily on average have lower monthly earnings, lower hours of work and a lower hourly wage. Of course these differences need not be related to speaking dialect itself but may be a reflection of other personal characteristics that correlate with dialect-speaking.

[4]Note that the graphs distinguish individuals who speak dialect daily from those who speak standard Dutch. However, those who speak standard Dutch might speak dialect once in a while or regularly but less than daily.

Table 2: Sample characteristics by gender and daily dialect-speaking

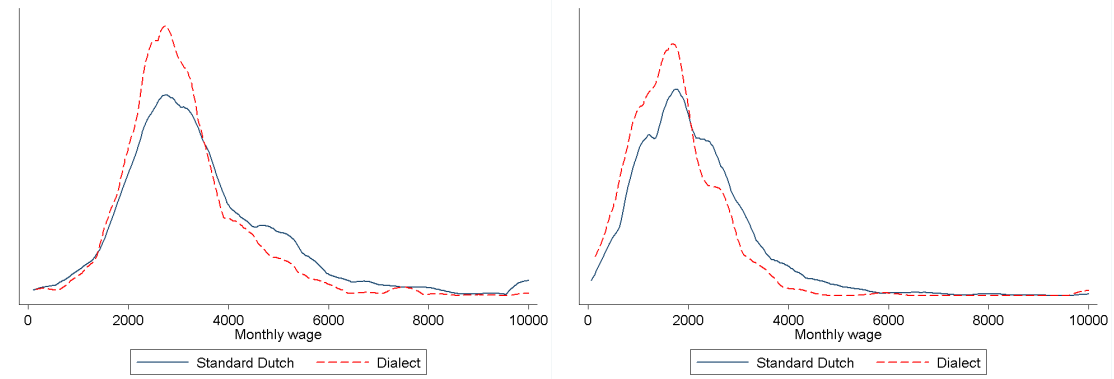
	Males		Females	
Dialect speakers	No	Yes	No	Yes
Speaks dialect (%)				
Never	68.0	0.0	71.7	0.0
Once in a while	21.3	0.0	19.0	0.0
Regularly	10.8	0.0	9.3	0.0
Daily	0.0	100.0	0.0	100.0
Personal characteristics				
Age	43.7	45.9	43.1	45.3
Education (%)				
Primary education	6.9	6.4	6.4	6.2
Lower secondary education	16.4	27.6	20.4	31.9
Intermediate secondary education	36.6	39.6	37.5	40.3
Higher education	40.2	26.4	35.7	21.5
Number of children	1.1	1.0	1.1	1.1
Living with a Partner (%)	79.1	79.6	77.7	79.9
Has a religion(%)	19.7	18.4	20.7	14.7
Urbanized area (%)	44.9	17.0	42.6	18.8
Province characteristics				
Log(GDP per capita)	10.5	10.4	10.5	10.4
Log(Employment)	6.9	6.5	6.9	6.4
Log(Population (1,000))	14.4	14.0	14.4	14.0
Area in use of main roads (km^2)	111.7	107.0	111.0	102.6
Distance to Amsterdam (5km)	14.3	23.7	14.7	23.6
N	8,154	2,265	10,240	1,922
Labor market and job characteristics				
Employment (%)	75.6	78.5	69.5	64.4
N	8,154	2,265	10,240	1,922
Monthly earnings (Euro)	3,482	3,078	2,120	1,736
N	5,581	1,468	5,913	990
Weekly working hours	39.8	40.0	28.5	26.9
N	5,084	1,406	5,633	999
Hourly wage (Euro)	20.9	18.6	17.5	15.5
N	4,613	1,236	4,989	843

Note: The level of education dummy variables are based on Statistics Netherlands categories, primary education, lower secondary education, intermediate secondary education and higher education. In an urbanized area population density is above 1,500 inhabitants per squared kilometer. N is the number of observations; n is the number of individuals.

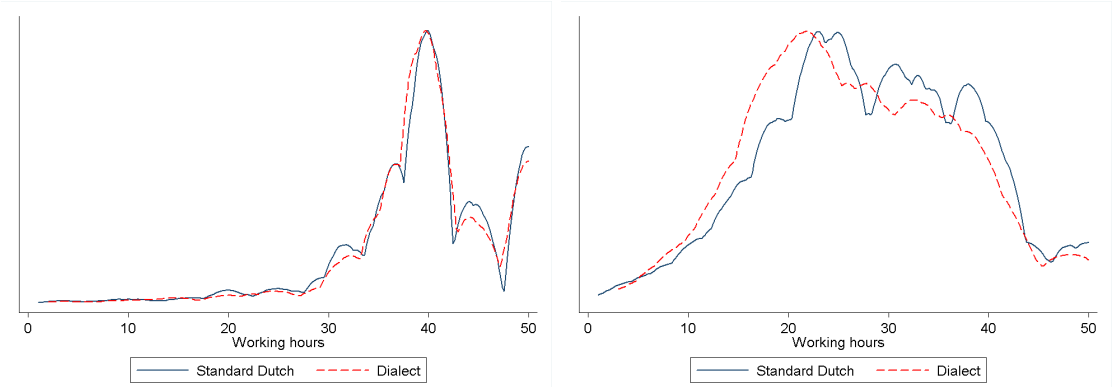
Figure 1: Kernel densities monthly earnings, working hours and hourly wages

Males Females

a. Monthly earnings



b. Weekly hours of work



c. Hourly wages

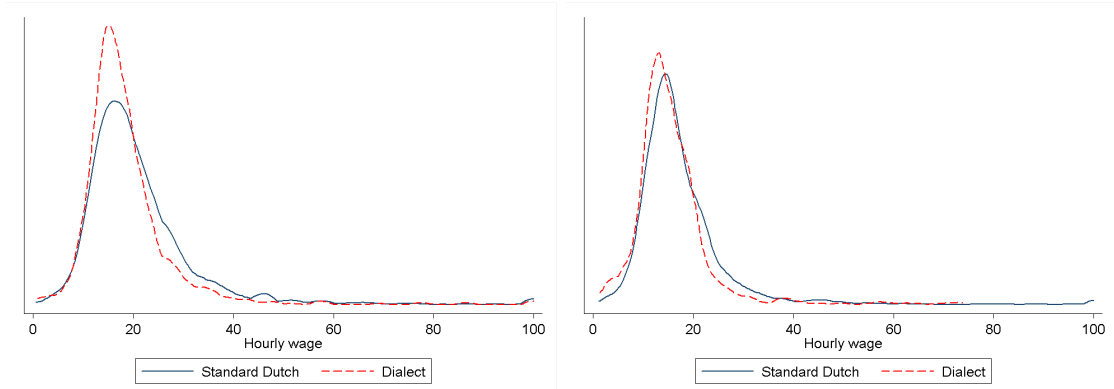
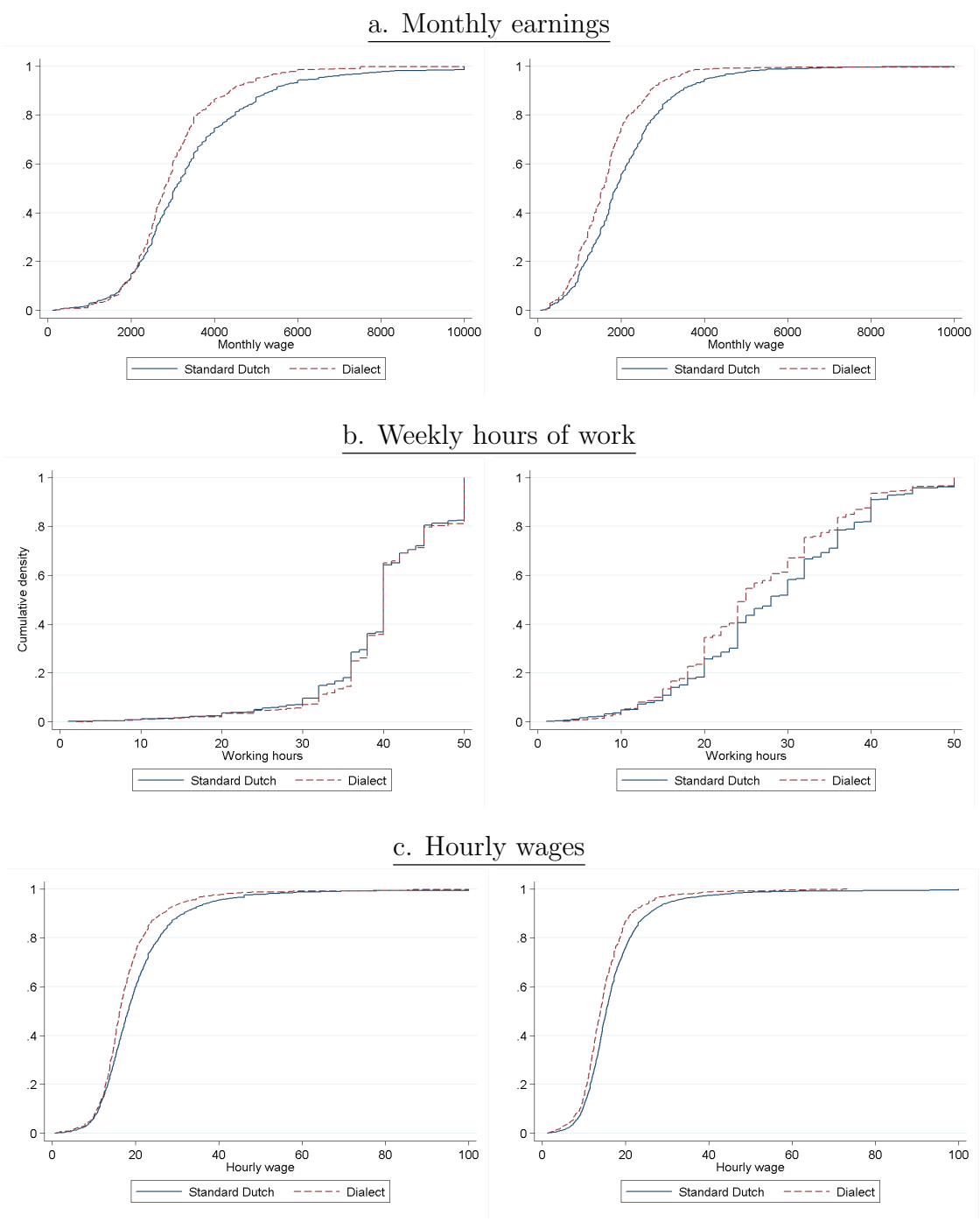


Figure 2: Cumulative distribution monthly earnings, working hours and hourly wages



3 Statistical model

3.1 Dialect and labor market performance

As discussed in the introduction, frequent usage of dialect can have causal effects on labor market performance for several reasons. First, frequent dialect-speaking may lead to worse command of the standard Dutch. Therefore frequent dialect-speaking may reduce productivity. Second, dialect-speakers may face discrimination in labor market. Dialects are sometimes labeled as lower-class languages. Although racial discrimination is not an issue in the Netherlands, bigoted employers may discriminate lower-class based on their speech pattern. Besides, dialects can signal workers' underlying skills.

However, there are several threats to identification of causal effects. First, labor market performance can reversely determine dialect usage. Workers in low-skilled occupations are not required to speak standard Dutch. Second, unobserved factors can account for lower productivity of dialect speakers. For example, less productive people may be more likely to stay in the place of birth and remain frequent dialect speakers. Third, there may be survey measurement errors in establishing the definition of dialects and whether or not individuals speak dialect frequently.

To investigate the causal effects of dialect usage, we first ignore these potential threats to identification and estimate equations by OLS:

$$Y_i = \alpha_1 + \gamma D_i + \beta_1 X_i + u_{1i} \tag{1}$$

where Y_i refers to one of the following labor market indicator: employment, monthly earnings, hours of work and hourly wage. Furthermore, D_i is a dummy variable for daily dialect-speaking, X_i is a vector of individual (gender, age, square age, native, having a partner, education level, number of children, having a religion and urbanization level of residence) and province characteristics (log of per capita GDP, log of population, area of main roads (km^2) and log of number of employment positions at each province).[5] To identify the causal effect of daily dialect-speaking, we need exclusion restrictions. We specify the determinants of speaking dialect daily as follows.

$$D_i = \alpha_2 + \theta Z_i + \beta_2 X_i + u_{2i} \tag{2}$$

[5]The provincial characteristics data are from Statistics Netherlands.

where Z_i is the excluded variable that affects dialect speaking but is exogenous to labor market outcomes. This is a model with continuous outcomes and binary endogenous variables. We identify the causal effects using a 2SLS estimation.

3.2 Identification

In order to account for potential endogeneity, we use an instrumental variable method. As standard Dutch originates from the Randstad area, people living closer to Randstad area are less likely to be dialect-speaking. We use the distance between the municipality of current residence and Amsterdam, the capital of the Netherlands as instrumental variable.[6] From estimating the determinants of daily dialect-speaking, we find that the distance to Amsterdam has a significant positive effect, after we control province variables even province fixed effects.

Our identifying assumption is that conditional on all observed characteristics, the distance to Amsterdam does not directly affect labor market outcomes of individual workers. To account for economic factors in one's place and province of residence, we control province characteristics such as province GDP, employment, roads area and population, as well as the urbanization level of neighborhood. Holding these economic factors fixed, geographic distance to Amsterdam is exogenously determined and it reflects the linguistic distance of local dialects to standard Dutch. So it does not affect labor market performance through other factors.

The argument above can be reinforced by the assumption that people do not move from the province of origin to provinces with a different dialect, or that at least internal migration across provinces is exogenous to labor market performance. A back-of-the-envelope calculation shows that this assumption is not too strong. According to Statistics Netherlands, the probability of migration between provinces is very low, 1.4-1.8% from 1995-2014. Suppose children do not move without parents[7]. So the province of living at 15 years old 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 where they lived at 15 years old is 72.6%. Moreover, many individuals may migrate between provinces using the same dialect, and some migrate back to province

[6]For reasons of confidentiality we only have information about the distance to Amsterdam in categorical measures of 5 km of distance, so that municipality of residence cannot be inferred from the distance.

[7]Migration without family is the major type of migration between provinces. The probability of migrating with family is 0.5-.6% from 1995-2014. The probability of staying in the province of birth is 93% at 15 years old.

of origin years later. Therefore, the actual staying rate is high for younger cohorts. In Appendix B, we also show that net migration inflow is uncorrelated with local economic situation, especially labor market factors at province level. Internal migration within the Netherlands does not seem to be a threat to the identification of a causal effect of daily dialect-speaking on labor market performance.

4 Parameter estimates

4.1 Baseline results

Table 3 presents OLS estimation results where the standard errors are clustered at individual level. In the first column only year fixed effects are included in addition to the dummy variable representing daily dialect speaking. The second column adds to that personal characteristics and the third column adds in addition province characteristics. The fourth column has province fixed effects replacing province characteristics.

Panel (a) of Table 3 shows the parameter estimates for the probability to be employed. Most specifications show a significant effect of dialect speaking. Apparently, the employment probability is not affected by speaking dialect daily when individual characteristics are included. Therefore, we do not consider the self-selection to employment with respect to dialect-speaking. Panel (b) shows the parameter estimates for three job characteristics: monthly earnings, weekly hours of work and hourly wages. Panel (b1) shows the parameter estimates for males, panel (b2) for females. In line with the cumulative distributions of Figure 2 the first column shows that dialect speaking has a negative effect for monthly earnings and hourly wages of males but not on the working hours of males. For females, dialect speaking has a negative effect on each of the job characteristics. When we subsequently introduce individual characteristics and province characteristics or province fixed effects many of the significant effects become smaller (in absolute size) and less significant. We only find persistent negative effects of dialect-speaking on monthly earnings of females and hourly wages of both genders, rather than working hours. This indicates that the penalty of dialect speaking for monthly earnings is mainly driven by hourly wage. Without controlling any for any other characteristic (except for calendar year fixed effects), dialect dialect speaking decreases males' hourly wage by 10.0% and females' hourly wage by 12.1%. As we include more characteristics and province characteristics, the effect of speaking dialect daily goes

Table 3: OLS parameter estimates effect of speaking dialect daily on labor market performance

	(1)	(2)	(3)	(4)	N
a. Employment					
1. Males	-0.014	-0.025	-0.017	-0.004	10,419
	(0.019)	(0.015)	(0.016)	(0.016)	
2. Females	-0.038*	-0.010	-0.009	-0.005	12,162
	(0.023)	(0.021)	(0.022)	(0.023)	
b. Job characteristics					
b1. Males					
Log monthly earnings	-0.093***	-0.050**	-0.026	-0.021	7,049
	(0.023)	(0.020)	(0.021)	(0.021)	
Log working hours	0.006	0.008	0.013	0.026*	6,490
	(0.013)	(0.013)	(0.013)	(0.013)	
Log hourly wage	-0.100***	-0.060***	-0.041*	-0.047*	5,849
	(0.024)	(0.021)	(0.022)	(0.022)	
b2. Females					
Log monthly earnings	-0.207***	-0.073**	-0.066*	-0.075*	6,903
	(0.039)	(0.034)	(0.035)	(0.039)	
Log working hours	-0.063**	0.007	0.005	-0.009	6,632
	(0.027)	(0.025)	(0.026)	(0.029)	
Log hourly wage	-0.121***	-0.061**	-0.051*	-0.041	5,832
	(0.028)	(0.026)	(0.028)	(0.030)	
Individual characteristics	N	Y	Y	Y	
Province characteristics	N	N	Y	N	
Province FE	N	N	N	Y	

Note: All estimates contain year fixed effects; results are reported with cluster standard errors at individual level. In Column (1), we regress labor market performance on speaking dialect without any control variables. In Column (2), we add individual characteristics and year fixed effects. In Column (3), we further control province characteristics. In Column (4), we control province fixed effects instead of province characteristics.

down. Speaking dialect daily decreases males wage by 4.1%, while this is 5.1% for females at 10% level.[8] When we compare columns (3) and (4), the difference in parameter estimates is small. This indicates that the province characteristics cover differences between provinces quite well such that province fixed effects do not add much.

Table 4 report the 2SLS estimates of the effect of daily dialect-speaking on labor market outcomes. As discussed before, we use the distance to Amsterdam as an instrumental variable for the probability that an individuals speaks dialect daily. Because the distance to Amsterdam is highly correlated with the province of residence we cannot use province fixed effects among the explanatory variables. For every estimate shown in Table 4 the F-test for the excluded variable in the first stage is very high indicating that our estimates do not suffer from a weak instrument.

Panel (a) shows our baseline parameter estimates. For males, the effects of dialect speaking on monthly earnings and hourly wages are quite strong and significantly negative. Monthly earnings are about 23% lower for dialect speakers while hourly wages are almost 15% lower.[9] Working hours are also lower for dialect speakers but the estimate is not significantly different from zero. For females monthly earnings and hourly wages are about 11-12% lower for dialect speakers but these effects are estimates with some imprecision so they do not differ from zero at conventional levels of significance.

In panel (b), we report the 2SLS parameter estimates from a random effects model. Our sample is an unbalanced panel data set. Since our main explanatory variable, the speaking daily dialect dummy does not vary much across years, we could not estimate individual fixed effects models and we estimated random effects model to account for unobserved heterogeneity. The differences between the parameter estimates in panels (a) and (b) are very small so our main conclusions are not affected by introducing random effects.

In panel (c) shows parameter estimates if we change the independent variable daily dialect-speaking to ever speaking-speaking, i.e. daily, regularly and once in a while. As was to be expected the magnitude of the effects for males become smaller but for both monthly earnings and hourly wages they are still negative and significant. Speaking dialect reduces male monthly earnings with 18.5% and reduces hourly wages with 12.8%. For females there is still no significant effect of dialect speaking on earnings and wages though the point estimates are substantial.

[8] Appendix A shows the parameter estimates for the hourly wage equation in more detail.

[9] Appendix A provides more detailed information about the 2SLS estimates of the log wage equation.

Table 4: 2SLS parameter estimates effect of speaking dialect daily on labor market performance

	Males				Females			
Variables	Effect speaking dialect daily		1st stage F-test	N	Effect speaking dialect daily		1st stage F-test	N
a. Baseline estimates								
Log monthly earnings	-0.226	(0.076)***	150.9	7,049	-0.112	(0.166)	75.2	6,903
Log working hours	-0.074	(0.046)	162.0	6,490	-0.016	(0.112)	94.8	6,632
Log hourly wage	-0.152	(0.073)**	145.2	5,849	-0.127	(0.119)	76.6	5,832
b. Random effects								
Log monthly earnings	-0.224	(0.046)***	246.3	7,049	-0.086	(0.082)	167.0	6,903
Log working hours	-0.074	(0.028)**	271.9	6,490	0.116	(0.090)	214.7	6,632
Log hourly wage	-0.171	(0.056)***	229.8	5,849	-0.127	(0.055)**	174.5	5,832
c. Speak dialect ever								
Log monthly earnings	-0.185	(0.061)***	249.9	7,049	-0.084	(0.123)	110.2	6,903
Log working hours	-0.062	(0.037)	254.9	6,490	-0.013	(0.087)	128.5	6,632
Log hourly wage	-0.128	(0.061)**	220.5	5,849	-0.099	(0.092)	100.6	5,832
d. Age < 40								
Log monthly earnings	-0.331	(0.101)**	60.4	2,178	-0.014	(0.228)	29.1	2,496
Log working hours	-0.086	(0.078)	64.2	1,907	0.028	(0.144)	32.6	2,359
Log hourly wage	-0.319	(0.114)***	59.2	1,686	-0.118	(0.149)	33.2	2,040
e. Higher education								
Log monthly earnings	-0.255	(0.113)**	56.7	3,094	0.101	(0.208)	28.6	2,943
Log working hours	0.037	(0.060)	59.8	2,867	-0.133	(0.145)	41.6	2,772
Log hourly wage	-0.275	(0.111)**	51.9	2621	0.087	(0.189)	33.5	2,493
f. Profession								
Professional ranking	-1.129	(0.408)***	155.0	6,990	-1.026	(0.483)**	94.2	6,997

Note: All results are reported with cluster standard errors at individual level. For all regressions, we use 2SLS estimation and include individual characteristics, calendar year fixed effects and province characteristics.

Panel (d) shows parameter estimates when we restrict our sample to the younger half of our sample i.e. workers up to 40 years old. As we discussed, younger cohorts have a higher probability to stay in the province of origin and geographic distance to Amsterdam is likely to be a valid instrumental variable. We find that the wage penalty is larger for younger male workers, while the effects on females remains insignificant. This suggests that the finding of wage penalty is not overestimated by self-selection of residence. If we estimate the same models on a sample of male workers older than 40 years there are no longer significant negative effects on hourly wages. For females we find no effects irrespective of their age. Therefore, the wage penalty of speaking dialects is more severe among younger cohorts.

4.2 What explains our findings?

Having established that there is a significant wage penalty of speaking dialect for males the question is how to explain this finding. To explore potential answers to this question Table 4 panels (e) to (f) present additional parameter estimates.

In panel (e) we limit the sample to workers with a high education. If speaking dialect is a potential signal of a lower productivity, we would expect the higher educated to suffer more from dialect speaking because their targeted occupations require more intellectual skills. The magnitude of the wage penalty of dialect speaking is larger for higher educated workers than for the average worker.

Finally, in panel (f) we show the effect of dialect speaking on professional ranking where we ranked professions from low to high on a scale from 1 to 9 according to the average male earnings in this profession (see Appendix C for details). For both males and females, dialect-speakers on average end up with occupations at lower professional rank.

5 Conclusions

In this paper we investigate whether speaking dialect affects labor market performance of individual workers. Using data from the Netherlands we conclude that speaking dialect daily does not affect the probability that an individual has a job but for males it does affect some job characteristics. Male workers who speak a dialect daily on average have significant lower monthly earnings and a significant lower hourly wage. For females we also find negative effect of dialect speaking on earnings and wages but these effects are not significantly different from

zero. Young and high educated male workers face a large wage penalty of dialect speaking. Finally, conditional on other personal characteristics including educational attainment we find the daily dialect-speakers are less likely to have a high-ranked profession.

Our data do not allow us to make a clear distinction between various mechanisms that lead to dialect-speaking having negative wage effects, in particular for male workers. It may be that dialect-speakers are perceived to have a lower productivity than workers who speak standard Dutch. Even if workers do not have difficulty in speaking standard Dutch at work it may still be obvious for colleagues and employer to recognize a dialect accent. Whether dialect-speakers actually have a lower productivity is hard to tell. We consider the fact that young and high educated workers face a larger wage penalty than their counterparts as an indication that perceived productivity is not in line with actual productivity. Mechanisms such as discrimination and signaling can explain the wage penalty better than actual productivity.

References

Bendick Jr., M., R. E. Rodriguez, and S. Jayaraman (2010). Employment discrimination in upscale restaurants: Evidence from matched pair testing. *Social Science Journal 47*, 802–818.

Budría, S. and P. Swedberg (2012). The impact of language proficiency on the impact of language proficiency on immigrants earnings in Spain. *IZA Discussion Paper No. 6957*.

Carlson, H. K. and M. A. McHenry (2006). Effect of accent and dialect on employability. *Journal of Employment Counseling 43*, 70–83.

Das, M. B. (2013). Exclusion and discrimination in the labor market. Background paper World Development Report.

Di Paolo, A. and J. L. Raymond (2012). Language knowledge and earnings in Catalonia. *Journal of Applied Economics 15*(1), 89–118.

Gao, W. and R. Smyth (2011). Economic returns to speaking 'standard Mandarin' among migrants in China's urban labour market. *Economics of Education Review 30*, 342–352.

Gorter, D. (2005). Three languages of instruction in Fryslan. *International Journal of the Sociology of Language 171*, 57–73.

Grogger, J. (2011). Speech patterns and racial wage inequality. *Journal of Human Resources 46*, 1–25.

Lang, K. (1986). A language theory of discrimination. *Quarterly Journal of Economics 101*, 363–382.

Miranda, A. and Y. Zhu (2013a). The causal effect of deficiency at English on female immigrants' labor market outcomes in the UK. IZA Discussion Paper No. 7841.

Miranda, A. and Y. Zhu (2013b). English deficiency and the native immigrant wage gap. *Economics Letters 118*(1), 38–41.

Smakman, D. (2012). The definition of the standard language: a survey in seven countries. *International Journal of the Sociology of Language 218*, 25–58.

Van Bezooijen, R. and W. Heeringa (2006). Intuitions on linguistic distance: geographically or linguistically based? In T. Koole, J. Nortier, and B. Tahitu (Eds.), *Vijfde sociolinguïstische conferentie*, pp. 77–87. Eburon Uitgeverij BV.

Yao, Y., A. Ohinata, and J. C. van Ours (2016). Educational consequences of language for young children. mimeo.

Yao, Y. and J. C. van Ours (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics 34*, 76–85.

Appendix A: Parameter estimates log wage equations

Table 5 shows the parameters of the wage equations obtained by OLS and the baseline 2SLS in more detail. In addition to the significant negative effects of daily dialect speaking age has a non-linear effect on hourly wages with for males an increase up to age 61 and a decrease later on. For females the maximum hourly wage is around age 55.[10] The number of children is not related to the wages of males but has a negative effect on the wages of females. For both males and females the highest wages are for the worker with the highest education. For males having a partner is positively associated with the wage, for female there is a negative association.

Apart for the wage penalty of dialect speaking which is much higher in the 2SLS estimates the effect of the other variables is very much the same in the 2SLS and the OLS estimates. Urbanization and province characteristics are not important as determinants of the hourly wages. The probability to speak a dialect daily – the first step of the 2SLS procedure – is higher in non-urban areas and in provinces with a low GDP per capita, a high employment and a low population with few main roads. And, of course, the distance to Amsterdam has a significant positive effect. The further away from Amsterdam the more likely an individual is to speak dialect. Finally, the probability to speak dialect is lower for males with a higher education.

Appendix B: Determinants of net migration inflow

Table 6 presents the correlation between internal migration and labor market factors at province level from 2010 to 2013. The data is from Statistics Nederlands. The dependent variable is net inflow of migrated population due to inter-municipality move, and the independent variables are GDP per capita, total compensation of employees (million euro), total number of employed people (per 1000), total working hours (million hours) and number of population (in 1000). We control calendar fixed effects and province fixed effects, the OLS results suggest that labor market factors (compensation, employed people and working hours) do not jointly affect migration inflow between provinces. Internal migration might be driven by unobserved location preference.

[10] The decline of the wage at a high age is due to the quadratic specification. This introduces a non-linearity in the age-wage profile but does not necessarily imply that such a decline is actually observed in the data.

Table 5: Parameter estimates of dialect-speaking effects on hourly wages

	OLS			2SLS – first step			Second step		
Males	Log wages			Dialect daily			Log wages		
Dialect daily	-0.047	(0.022)	**				-0.152	(0.073)	**
Age	0.059	(0.008)	***	0.009	(0.006)		0.060	(0.008)	***
Age-squared/100	-0.048	(0.009)	***	-0.010	(0.007)		-0.050	(0.009)	***
Lower second educ	0.034	(0.055)		0.029	(0.056)		0.040	(0.055)	
Intermediate sec edu	0.125	(0.053)	**	-0.063	(0.053)		0.120	(0.053)	**
Higher education	0.402	(0.054)	***	-0.106	(0.053)	**	0.392	(0.053)	***
Number of children	-0.008	(0.009)		-0.013	(0.008)		-0.009	(0.009)	
Partner	0.063	(0.024)	***	-0.029	(0.023)		0.058	(0.024)	**
Very urban	0.006	(0.032)		-0.035	(0.020)	*	0.010	(0.031)	
Moderately urban	0.056	(0.036)		0.039	(0.023)	*	0.064	(0.036)	*
Slightly urban	0.062	(0.036)	*	0.122	(0.027)	***	0.079	(0.039)	**
Not urban	0.021	(0.039)		0.190	(0.034)	***	0.053	(0.044)	
Religious	-0.001	(0.020)		0.000	(0.018)		-0.000	(0.020)	
Log GDP per capita				-0.3407	(0.100)	***	-0.126	(0.093)	
Log Employment				0.928	(0.187)	***	-0.005	(0.231)	
Log Population				-0.876	(0.181)	***	0.106	(0.236)	
Area Main Roads				-0.001	(0.000)	***	-0.001	(0.000)	**
Distance to Amsterdam				0.019	(0.002)	***			
Females									
Dialect daily	-0.041	(0.030)					-0.127	(0.118)	
Age	0.044	(0.007)	***	-0.002	(0.005)		0.043	(0.007)	***
Age-squared/100	-0.040	(0.008)	***	0.003	(0.006)		-0.040	(0.008)	***
Lower second educ	-0.213	(0.075)	***	0.049	(0.043)		-0.204	(0.076)	***
Intermediate sec edu	-0.094	(0.073)		0.012	(0.040)		-0.090	(0.073)	
Higher education	0.153	(0.074)	**	-0.038	(0.039)		0.155	(0.073)	**
Number of children	-0.027	(0.009)	***	0.007	(0.008)		-0.027	(0.009)	***
Partner	-0.060	(0.023)	***	0.008	(0.018)		-0.058	(0.023)	**
Very urban	-0.030	(0.033)		-0.020	(0.018)		-0.036	(0.033)	
Moderately urban	-0.050	(0.037)		-0.007	(0.021)		-0.054	(0.037)	
Slightly urban	-0.068	(0.038)	*	0.015	(0.021)		-0.063	(0.038)	*
Not urban	-0.053	(0.040)		0.131	(0.030)	***	-0.039	(0.045)	
Religious	0.028	(0.021)		-0.010	(0.014)		0.027	(0.021)	
Log GDP per capita				-0.632	(0.074)	***	-0.047	(0.130)	
Log Employment				1.032	(0.154)	***	0.124	(0.243)	
Log Population				-0.918	(0.149)	***	-0.094	(0.243)	
Area Main Roads				-0.002	(0.000)	***	-0.000	(0.000)	
Distance to Amsterdam				0.014	(0.002)	***			

Note: All estimates contain calendar year fixed effects; the OLS estimates also contain province fixed effects. The main parameter estimates are also reported in Table 3 panel b and Table 4 panel a.

Table 6: Determinants of net migration inflow

	Net inflow	
GDP per capita	0.025	(0.116)
Total compensation	0.396	(0.566)
Number of employees	-7.540	(92.098)
Total working hours	-14.67	(51.806)
Population	-5.385	(32.009)
Year fixed effects	Yes	
Province fixed effects	Yes	
F-statistics for wage factors	0.41	
Adjusted R^2	0.91	
# Obs.	48	

Appendix C: Details on the data

The professional ranking used in Table 4 panel f is based on the average monthly earnings. The ranking is as follows:

Profession	Monthly earnings (Euro)	Professional ranking
Unskilled and trained manual work	1623	1
Semi-skilled manual work	2447	2
Agrarian profession	2562	3
Skilled and supervisory manual work	2646	4
Other mental work	2865	5
Intermediate academic or independent	3164	6
Intermediate supervisory or commercial	3335	7
Higher academic or independent	4160	8
Higher supervisory profession	5230	9
Average	3398	5