DISCUSSION PAPER SERIES

The Wage Penalty of Dialect-Speaking

Yuxin Yao
Jan C. van Ours

October 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Wage Penalty of Dialect-Speaking

Yuxin Yao
CentER, Tilburg University

Jan C. van Ours
Erasmus University Rotterdam,
Tinbergen Institute, University of Melbourne, CEPR and IZA

ABSTRACT

The Wage Penalty of Dialect-Speaking[*]

Our paper studies the effects of dialect-speaking on job characteristics of Dutch workers, in particular on their hourly wages. The unconditional difference in median hourly wages between standard Dutch speakers and dialect speakers is about 10.6% for males and 6.7% for females. If we take into account differences in personal characteristics and province fixed effects male dialect speakers earn 4.1% less while for females this is 2.8%. Using the geographic distance to Amsterdam as an instrumental variable to dialect-speaking, we find that male workers who speak a dialect earn 11.6% less while for female workers this is 1.6%. Our main conclusion is that for male workers there is a significant wage penalty of dialect-speaking while for female workers there is no significant difference.

Corresponding author:

Jan C. van Ours
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
E-mail: vanours@ese.eur.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. Though among linguists there is no common definition of dialects, a dialect is usually referred to as a variation of a language used by a particular group. A dialect may associate with 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]

However, as in most cases, we refer to a dialect as a regional speech pattern. A dialect is a variation of the standard language, used in limited regions and different in mainly pronunciation, and sometimes vocabulary and grammar. Dialects can be acquired without training and play a role in informal communication, while the standard language is the instruction medium at schools. Speaking with a local dialect accent may reflect lower language ability or lack of communication with people from other regions. Therefore, it is of interest to explore how dialect speech patterns affect labor market performance.

To study the effects of speech patterns, Grogger (2011, 2014) used NLSY data in combination with audio-information about how individuals speak. 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. There is also a wage penalty of perceived Southern speech pattern. The origin of the effects of speech patterns on wages is not clear. Grogger (2011) suggests two possible explanations. First, a non-standard speech pattern reduces workplace productivity. Second, there can be a causal effect working

---

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

through discrimination from bigoted employers. 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. Language speech differences among workers may increase production costs (Lang, 1986) or language speech may influence the severity of discrimination by employers. A second explanation is that speech is a (negative) signal of unobserved productivity.

Our paper studies the relationship between dialect-speaking and labor market performance, in particular hourly wages. We study the Netherlands as an example of a country with a lot of commuting such that spatial segregation is limited. This is not only because there are various dialects spoken, but also Dutch natives are more homogeneous in terms of culture, physical characteristics and economic wealth than natives from larger countries. Moreover, to compare native dialect speakers with standard Dutch speakers, we can obtain purer effects of speech pattern than comparing immigrants with natives. To deal with endogeneity of dialect-speaking behavior, we use geographic distance to Amsterdam as an instrumental variable. Dialect-speaking is more prevalent in municipalities located far from Amsterdam. Our main finding is that there is a significant wage penalty of dialect-speaking for males while for females dialect-speaking does not affect wages.

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

2 Data and Background

2.1 Linguistic Background

The predominantly spoken language of the Netherlands is Standard Dutch, originating from the urban areas of Noord-Holland. Besides Standard Dutch, the regional languages and dialects spoken in the Netherlands are remarkably diverse, including Frisian, Limburgish, and Low Saxon. Frisian, mostly spoken in the province of Friesland, is recognized as a separate language and promoted by the local government. In Friesland both Standard Dutch and Frisian are considered official languages and instruction media at school. More than 80% of the adult inhabitants understand verbal Frisian, but only a small minority can write the language (Gorter, 2005). In our paper we refer to Frisian as a dialect for simplicity. Other regional languages include Limburgish and Low Saxon, which enjoy the status as "official regional languages" in related regions although there is no clear regulation regarding government support. Limburgish is spoken in the province of Limburg by about 75% of the inhabitants and Low Saxon is spoken in the provinces of Groningen, Drenthe, Overijssel and Gelderland by approximately 60% of the inhabitants. Other provinces also have dialects such as Brabantish, spoken in Noord-Brabant or Zeelandic in Zeeland (see an overview in Driessen (2005) and Cheshire et al. (1989)).

Distances between languages depend on characteristics such as vocabulary, pronunciation, syntax and grammar. To quantify distances between languages various methods are used. Levenshtein (1966) proposed an algorithm based on the minimum number of steps to change a particular word in one language to the same word in a different language. The overall distance between two languages is based on the average difference for a list of words for which often but not always the 100 words from Swadesh (1952) are used. Levenshtein's method can be based on written words but can also be based on phonetic similarities. This is especially helpful when comparing dialects as often these are spoken but not used in writing. Table 1 provides a dialect indicator at the province level. Van Bezooijen and Heeringa (2006) use two samples of Dutch dialects and apply the Levenshtein distance measure to calculate

Table 1: Dialect-speaking by province

Province	Linguistic distance	Distance to Amsterdam	Speaking dialect (%)				N
			Daily	Regularly	Sometimes	Never	
Drenthe	19	129	34	12	17	37	655
Flevoland	12	39	1	5	11	83	495
Friesland	37	106	48	9	13	30	1,110
Gelderland	28	82	14	9	20	57	3,106
Groningen	28	161	22	12	20	46	904
Limburg	32	160	68	7	10	15	1,628
Noord-Brabant	28	96	22	15	27	36	3,983
Noord-Holland	12	21	3	2	8	87	2,940
Overijssel	29	111	25	15	28	32	1,545
Utrecht	18	39	3	4	11	82	1,594
Zeeland	29	129	29	16	26	29	489
Zuid-Holland	12	52	3	2	10	85	4,127
Total	23	81	17	8	16	57	22,576

Note: Distance to Amsterdam in kilometers. Averaged over the individuals in our sample.
Source linguistic distance: Van Bezooijen and Heeringa (2006).

the average linguistic distances between provincial dialects and standard Dutch. In our paper, we use their distances, which are based on the New Dialect Sample. These distance measures are calculated from 100 words. 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.

Table 1 also shows that the geographical distance to Amsterdam of the individuals in our sample is on average the largest in Gelderland and Groningen and the smallest in Noord-Holland where Amsterdam is located. Dialect-speaking is most prevalent in Limburg where 68% of the individuals in our sample 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 the sample indicating that they speak dialect daily. In addition, Noord-Holland and Zuid-Holland are the provinces with the highest share of individuals in our sample who never speak dialect.

Figure 1 provides a graphical representation on the relationship between dialect characteristics and the geographical distance to Amsterdam at the provincial level. Figure 1a shows that linguistic distance and geographical distance at highly correlated with Friesland and

Figure 1: Linguistic distance, percentage dialect speakers and geographical distance to Amsterdam by province

a. Linguistic distance   b. Percentage dialect speakers
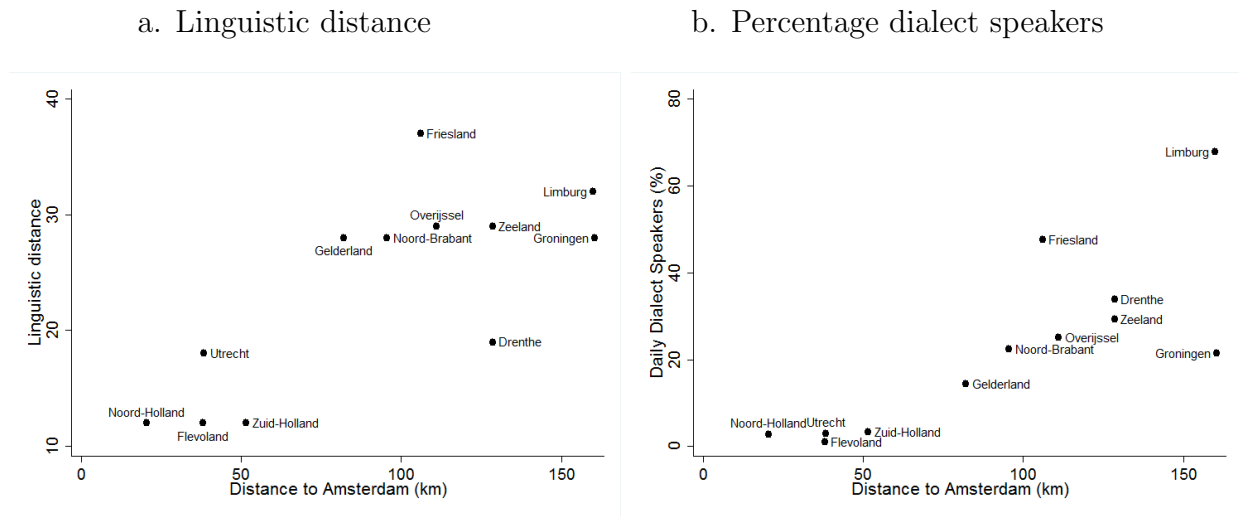
Source: see Table 1.

Drenthe as outliers. In Drenthe, the linguistic distance to standard Dutch is smaller than in other provinces with the same geographical distance to Amsterdam while in Friesland the linguistic distance to standard Dutch is larger than it is in comparable provinces. Figure 1b shows that there is also a strong correlation between the share of (daily) dialect speakers and the distance to Amsterdam, with Friesland and Limburg as outliers.

2.2 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 (see for details: www.lissdata.nl). We use the first seven waves of panel data from 2008 to 2014 initially focusing 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. Based on average weekly working hours and personal monthly gross earnings, we calculated hourly wages.

As is usual in studies on language effects, we rely on self-reported information. Respondents indicate dialect usage by answering the questions: *Do you ever speak dialect?* Respondents can choose from *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[2] To focus on the effects of dialect speaking pattern and exclude the effect of language deficiency, we remove the sample of 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 2 shows the summary statistics by gender and daily dialect-speaking. Comparing dialect speakers with Dutch speakers, we will find that dialect speakers have a slightly lower education and are less likely to live in urbanized areas. We do not find much difference in province characteristics between the two groups. If we look at labor market characteristics, employment rate and weekly working hours are similar between dialect speakers and Dutch speakers. However, dialect speakers on average have lower hourly wages and lower monthly earnings.

In 2008 the minimum wage in the Netherlands was about 7.8 Euro while in 2014 it was about 8.5 Euro. In the original sample some of the hourly wages are far below the minimum wage. To avoid a bias in the parameter estimates we removed all observations with an hourly wage below 7.5 Euro from the sample[3]. The densities and cumulative distributions of hourly wages by dialect status are presented in Figure 2. The differences between males and females are not big but individuals who speak dialect daily on average have lower hourly wages. The median wage of standard Dutch speaking males is 18.75 Euro while for standard Dutch speaking females this is 15.86 Euro. Among dialect-speakers the median hourly wage is 16.76 Euro for males and 14.79 Euro for females. So, on average for males dialect speakers

---

[2]Since Frisian is an official language rather than a dialect, we refer to the survey question "*which language do you generally speak at home?*" for respondents from Friesland.

[3]2.6% of the total observations are removed.

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

	Males		Females	
Dialect speakers	No	Yes	No	Yes
Speaks dialect (%)				
Never	68	0	72	0
Once in a while	21	0	19	0
Regularly	11	0	9	0
Daily	0	100	0	100
Personal characteristics				
Age	44	46	43	45
Education (%)				
Primary education	7	6	6	6
Lower secondary education	16	28	20	32
Intermediate secondary education	37	40	37	40
Higher education	40	27	36	22
Number of children	1.1	1.0	1.1	1.1
Living with a Partner (%)	79	80	78	80
Has a religion(%)	20	18	21	15
Urbanized area (%)	45	17	43	19
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)	112	107	111	103
Distance to Amsterdam (km)	70	120	75	120
N	8,154	2,265	10,240	1,922
Labor market and job characteristics				
Employment (%)	76	78	70	64
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.

Figure 2: Kernel densities and cumulative distribution of hourly wages

a. Kernel densities

b. Cumulative densities
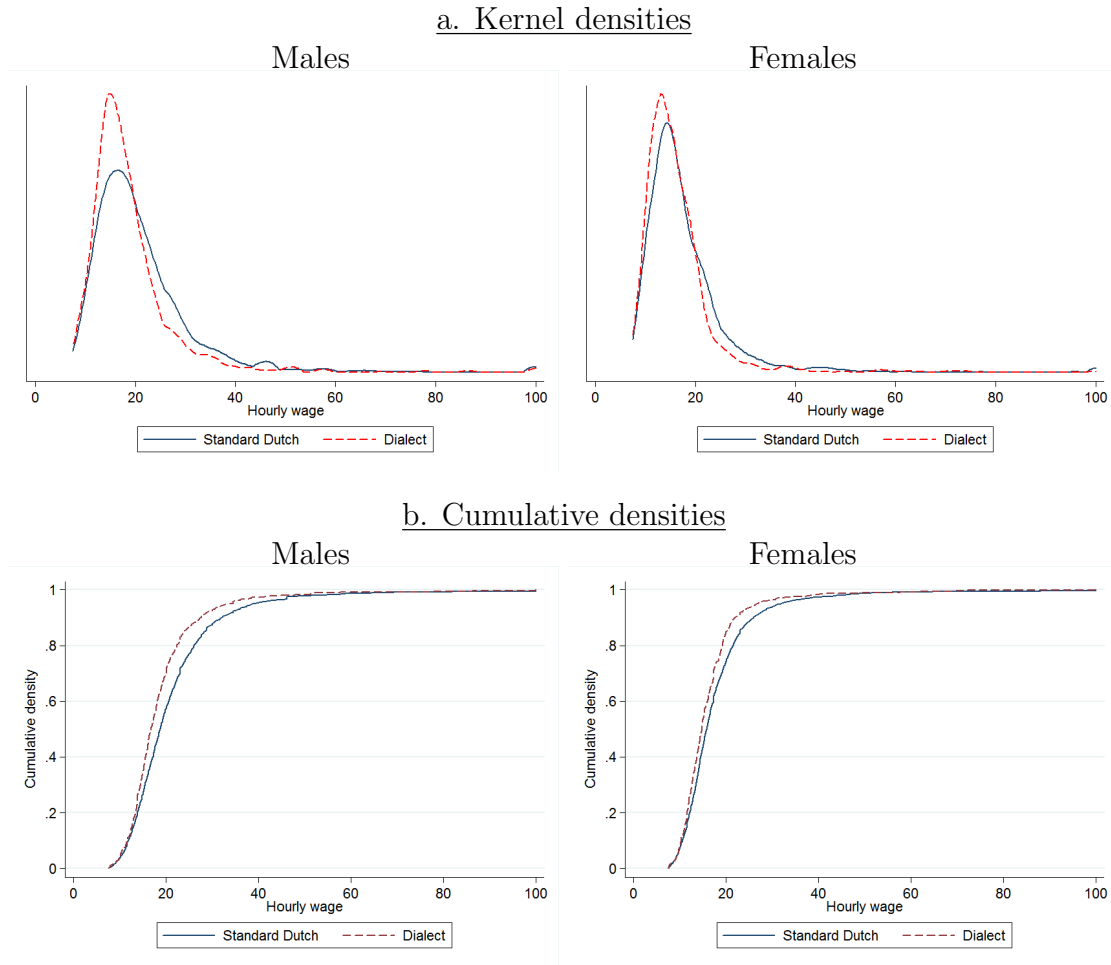
earn 10.6% less than non-dialect speakers. For females, the difference is 6.7%. Of course, these differences need not be related to speaking dialect itself but may be explained by other personal characteristics that correlate with dialect-speaking.

3 Statistical model

3.1 Dialect and labor market performance

As discussed in the introduction, frequent dialect-speaking can have a causal effect on labor market performance for several reasons. First, frequent dialect-speaking may lead to worse

command of the standard Dutch and reduce workplace productivity. Second, dialect-speakers may face discrimination in the labor market. Dialects are sometimes labeled as speech patterns of lower social classes. Prejudiced employers may discriminate against non-standard speech pattern. Finally, a speech pattern can be a negative signal of workers' underlying but unobserved skills.

There are several threats to identification of a causal effect. First, labor market performance can reversely determine dialect use. Workers in low-skilled occupations for example are not required to speak standard Dutch. Second, unobserved factors can account for a 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 initially refers to one of the following labor market indicators: employment, monthly earnings, hours of work, hourly wages. 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).[4] To identify the causal effect of daily dialect-speaking, we need exclusion restrictions. We specify the determinants of speaking dialect daily as follows.

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

---

[4]The provincial characteristics data are from Statistics Netherlands; see for further details Appendix A.

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.

3.2 Identification

In order to account for potential endogeneity, we use an instrumental variable method. As standard Dutch originates from the province Noord-Holland, people living closer to this province are less likely to speaking dialects. We use the distance between the municipality of current residence and Amsterdam, the capital of the Netherlands as instrumental variable.[5] We find that the distance to Amsterdam has a significant positive effect on the probability of daily dialect-speaking, after we control for province characteristics.

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 the 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 identifying assumption implies that people do not move away from the place of origin, or at least that internal migration 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% annually from 1995-2014. Suppose children and adolescents never move without parents[6], the province of residence is very likely to be the province of origin at least for younger cohorts. For instance, for a 35-year-old in 2014, the probability of staying in the province of origin

---

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

[6]Children and adolescents who move to other provinces should be with family. The probability of migrating with family is as low as 0.5-0.6% annually, while migration without family is the major type of migration in the Netherlands. For instance, the probability of staying in the province of birth is as high as 93% for a 15-year-old.

at 15 years old is 72.6%. Moreover, many individuals may move between provinces several times during lifetime or even migrate back to province of origin, while others do not move at all. And when we focus on Dutch natives, the actual staying rate is even higher than our back-of-the-envelope calculation result. In addition, to reinforce the argument of exogenous migration, we also show that net migration inflow is uncorrelated with local economic factors at the province level (See details in Appendix B). Finally, given that migration between provinces is not endogenous to labor market performance, migration between municipalities within a province is of less concern to identification. This is because commuting from home to work is common between municipalities, and people are not very motivated to move within a province. Therefore, 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 parameter estimates where the standard errors are clustered at the level of the individual. In the first column only calendar year fixed effects and individual characteristics are included in addition to the dummy variable representing daily dialect-speaking. The second column adds province characteristics to that, and column (3) we replaced province characteristics by province fixed effects.

Panel (a) of Table 3 shows the parameter estimates for males, panel (b) does the same for females. The first rows concern the probability to be employed. Neither for males nor for females, there is a significant effect of dialect speaking. Apparently, the employment probability is not affected by speaking dialect daily irrespective of whether province variables are included. The second rows show the estimates on monthly earnings. If province characteristics or fixed effects are not taken into account there are significant negative effects of dialect-speaking. If province characteristics are included for males the effect of

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

	(1)		(2)		(3)		N
a. Males							
1. Employment	-0.019	(0.016)	-0.013	(0.016)	-0.000	(0.016)	10,419
2. Log monthly earnings	-0.044	(0.019)**	-0.024	(0.019)	-0.019	(0.019)	6,921
3. Log working hours	0.008	(0.013)	0.012	(0.013)	0.024	(0.014)	6,362
4. Log hourly wage	-0.054	(0.019)***	-0.037	(0.019)*	-0.041	(0.020)**	5,721
5. Professional ranking	-0.410	(0.104)***	-0.368	(0.107)***	-0.328	(0.112)***	6,862
6. Non-manual job	-0.085	(0.024)***	-0.079	(0.248)***	-0.075	(0.027)***	6,852
7. Happiness	0.034	(0.057)	0.045	(0.060)	0.080	(0.061)	5,273
b. Females							
1. Employment	-0.013	(0.021)	-0.013	(0.022)	-0.011	(0.023)	12,162
2. Log monthly earnings	-0.059	(0.029)**	-0.056	(0.031)**	-0.085	(0.033)***	6,668
3. Log working hours	-0.002	(0.024)	-0.005	(0.026)	-0.030	(0.028)	6,397
4. Log hourly wage	-0.032	(0.020)	-0.024	(0.021)	-0.028	(0.023)	5,597
5. Professional ranking	-0.385	(0.103)***	-0.402	(0.108)***	-0.425	(0.120)***	6,763
6. Non-manual job	-0.074	(0.024)***	-0.078	(0.025)***	-0.098	(0.028)***	6,763
7. Happiness	0.018	(0.069)	-0.007	(0.071)	0.028	(0.074)	5,072
Province characteristics	No		Yes		No		
Province fixed effects	No		No		Yes		

Note: All estimates contain individual characteristics and year fixed effects; results are reported with cluster standard errors at individual level. In column (2) we also include province characteristics, while province fixed effects are included in column (3).

dialect-speaking on monthly earnings are no longer significantly different from zero. For females there is still a significant negative effect. Weekly working hours are not influenced by dialect-speaking for both males and females. Finally, as shown in the fourth rows, there is a significant negative effect of daily dialect-speaking on hourly wages of males of 4.1% when province fixed effects are included. For females, the negative effect is 2.8%, but this effect is not different from zero at conventional levels of significance. For all the regressions, the results are robust regardless of whether we add province characteristics or province fixed effects.

The fifth rows show the effect of dialect-speaking on the type of jobs, ranking professions from low to high on a scale from 1 to 9 according to the average male earnings in the profession (see Appendix A for details). Clearly, dialect-speaking lowers the rank of the profession. This is also the case if we distinguish between manual and non-manual job. Workers who speak dialect daily are less likely to work in a non-manual job. Finally, the seventh rows show the effects of dialect-speaking on happiness. The happiness indicator is based on self-reported answers to the question "*On the whole, how happy would you say you are?*" Happiness is measured on an ordinal scale from 0 (very unhappy) to 10 (very happy). As is shown, for both males and females happiness is positively though not significantly affects by dialect-speaking.[7]

So, for males the main effect of dialect-speaking seems to be that hourly wages are lower. For females there is also an effect on working hours which jointly with the effect on hourly wages leads to a significant effect on monthly earnings. However, since there is a choice of working hours while the hourly wage is much more difficult to be influenced by individual choice we focus the effects of dialect-speaking on hourly wages.

The first column of Table 4 reports the parameter estimates for hourly wages in detail. For both males (panel a) and females (panel b), age is positively related to hourly wages. Workers age 55-plus earn substantially more than workers aged 45 to 54, who in their turn

---

[7]Apart from including the explanatory variables in the other equations, the happiness estimates also include log monthly earnings as an explanatory variable. For both males and females higher earnings on average imply a higher level of happiness.

Table 4: Parameter estimates of dialect-speaking effects on log hourly wages

	OLS Log wages			2SLS: first stage Dialect daily			2SLS: Second stage Log wages		
a. Males									
Dialect daily	-0.037	(0.019)	*				-0.116	(0.066)	*
Age 25-34	0.162	(0.054)	***	0.047	(0.046)		0.167	(0.054)	***
Age 35-44	0.379	(0.053)	***	0.044	(0.046)		0.383	(0.054)	***
Age 45-54	0.488	(0.054)	***	0.074	(0.047)		0.492	(0.054)	***
Age 55+	0.606	(0.055)	***	0.051	(0.047)		0.610	(0.058)	***
Lower secondary educ.	0.031	(0.047)		0.031	(0.058)		0.036	(0.048)	
Intermediate sec educ.	0.126	(0.045)	***	-0.065	(0.054)		0.125	(0.046)	**
Higher education	0.388	(0.046)	***	-0.106	(0.054)	**	0.385	(0.047)	***
Number of children	0.002	(0.008)		-0.010	(0.009)		0.001	(0.008)	
Partner	0.054	(0.021)	**	-0.033	(0.023)		0.050	(0.022)	**
Very urban	-0.013	(0.028)		-0.034	(0.021)	*	-0.008	(0.029)	
Moderately urban	0.042	(0.031)		0.039	(0.023)	*	0.045	(0.031)	
Slightly urban	0.041	(0.033)		0.116	(0.028)	***	0.053	(0.034)	
Not urban	-0.006	(0.034)		0.193	(0.035)	***	0.009	(0.037)	
Religious	-0.007	(0.018)		-0.001	(0.019)		-0.007	(0.018)	
Log GDP per capita	-0.042	(0.073)		-0.400	(0.103)	***	-0.069	(0.078)	
Log Employment	-0.091	(0.194)		0.910	(0.190)	***	-0.071	(0.020)	
Log Population	0.191	(0.199)		-0.860	(0.184)	***	0.161	(0.202)	
Area Main Roads/100	-0.001	(0.000)		-0.130	(0.046)	***	-0.096	(0.036)	
Distance to Amsterdam/10				0.187	(0.016)	***			
F-test excluded instrument				140.2					
b. Females									
Dialect daily	-0.024	(0.021)					-0.016	(0.098)	
Age 25-34	0.162	(0.032)	***	0.019	(0.034)		0.162	(0.031)	***
Age 35-44	0.290	(0.033)	***	-0.007	(0.035)		0.291	(0.033)	***
Age 45-54	0.312	(0.032)	***	0.028	(0.036)		0.313	(0.033)	***
Age 55+	0.393	(0.036)	***	0.019	(0.036)		0.393	(0.036)	***
Lower secondary educ.	-0.175	(0.072)	**	0.050	(0.040)		-0.176	(0.072)	***
Intermediate sec. educ.	-0.078	(0.070)		0.019	(0.041)		-0.078	(0.070)	
Higher education	0.151	(0.071)	**	-0.032	(0.040)		0.151	(0.070)	**
Number of children	-0.003	(0.008)		0.001	(0.008)		-0.003	(0.008)	
Partner	-0.059	(0.020)	***	0.005	(0.019)		-0.059	(0.020)	***
Very urban	-0.046	(0.029)		-0.017	(0.018)		-0.045	(0.030)	
Moderately urban	-0.047	(0.032)		-0.003	(0.021)		-0.047	(0.032)	
Slightly urban	-0.063	(0.033)	**	0.018	(0.030)		-0.062	(0.045)	*
Not urban	-0.044	(0.034)		0.119	(0.030)	***	-0.045	(0.050)	
Religious	0.028	(0.018)		-0.009	(0.014)		0.028	(0.018)	
Log GDP per capita	0.030	(0.089)		-0.656	(0.068)	***	0.035	(0.114)	
Log Employment	0.021	(0.201)		1.051	(0.148)	***	0.017	(0.210)	
Log Population	-0.003	(0.201)		-0.928	(0.143)	***	0.017	(0.209)	
Area Main Roads/100	-0.000	(0.000)		0.247	(0.045)	***	-0.021	(0.038)	
Distance to Amsterdam/10				0.146	(0.017)	***			
F-test excluded instrument				75.0					

Note: 5,721 observations for males, 5,597 observations for females; all estimates contain calendar year fixed effects and province fixed effects. Professional ranking ranges from 0 to 9. Happiness ranges from 0 to 10. The main parameter estimates are also reported in Table 3 Panels a4 and b4 and Table 5 Panel a1.

15

earn more than workers ages 35 to 44, et cetera. For males the age effect is stronger than for females. Educational attainment also has a positive effect on hourly wages, although again this effect is stronger for males than for females. The main difference between males and females is the effect of the partner. While males with a partner earn more than males without a partner, for females this is the other way around. Of course, this is not necessarily a causal effect, i.e. it could even be reverse causality with women with higher earnings being less likely to have a partner or men with higher earnings being more likely to have partners. Religion and urbanization level of residence does not have significant effects on hourly wages. The latter is no surprise as many individuals work in a different municipality from where they live.

The second column shows the determinants of daily dialect-speaking. The main personal characteristics influencing dialect-speaking are degree of urbanization of the residence. In rural municipalities the probability to speak a dialect daily is substantial higher than in urban municipalities. As discussed before, we use the distance to Amsterdam as an instrumental variable for the probability that an individuals speaks dialect daily. As shown, for both males and females this variable is highly significant. The parameter estimates suggests that living geographically closer to Amsterdam is associated with a higher probability of speaking dialects by 1-2 percentage points. Because the distance to Amsterdam is correlated with the province of residence we cannot use province fixed effects among the explanatory variables. Provincial characteristics seem to matter too but this is probably related to the geographical location of the provinces (as indicated in Figure 1).

We use the estimates in column 2 as the first-stage for the 2SLS-estimates presented in column 3. The F-statistics for the excluded variable in the first stage are very high indicating that our estimates do not suffer from a weak instrument. Relying on a 2SLS method, dialect-speaking decreases males' hourly wages by 11.6% at 10% significant level and decreases females' hourly wages by 1.6% though the latter effect does not differ significantly from zero.

Apart for the wage penalty of dialect speaking which for males is much higher in the 2SLS estimates the effect of other variables are very much the same in the 2SLS and the

16

OLS estimates. Urbanization and province characteristics are not important as determinants of the hourly wages. For females, the wage penalty of dialect-speaking is smaller in the 2SLS-estimates. And, as before the effect is insignificantly different from zero.

Table 5: Parameter estimates effects of dialect-speaking; sensitivity analysis

	Males				Females			
	Effect dialect speaking		1st stage F-test	N	Effect dialect speaking		1st stage F-test	N
a. OLS								
1. Baseline estimates	-0.041	(0.020)**		5,721	-0.028	(0.023)		5,597
2. No Friesland	-0.054	(0.020)**		5,642	-0.024	(0.024)		5,350
3. No Randstad	-0.038	(0.021)*		3,361	-0.032	(0.024)		3,188
b. 2SLS								
1. Baseline estimates	-0.116	(0.066)*	140.2	5,721	-0.016	(0.098)	75.0	5,597
2. Ever dialect speaking	-0.097	(0.056)*	214.8	5,721	-0.013	(0.075)	100.2	5,597
3. Professional ranking	-1.007	(0.407)**	150.4	6,862	-0.926	(0.484)*	95.0	6,763
4. Non-manual job	-0.166	(0.086)*	150.4	6,862	-0.088	(0.090)	95.0	6,763
5. Age less than 40	-0.190	(0.100)*	65.5	1616	-0.056	(0.126)	36.2	1,953
6. Age 40 or more	-0.083	(0.081)	97.1	5,105	0.001	(0.127)	46.2	3,644
7. High educated	-0.242	(0.104)**	51.9	2,587	0.106	(0.173)	29.8	2,445
8. Non-high educated	-0.042	(0.085)	91.8	3134	-0.072	(0.114)	49.6	3,152

Note: Randstad is assumed to consist of the provinces Noord-Holland, Flevoland, Utrecht, Zuid-Holland. All results are reported with cluster standard errors at individual level. The values of the interaction with linguistic distance are divided by 10 and then range from 0 to 3.7. Professional ranking ranges from 0 to 9. All regressions include individual characteristics, calendar year fixed effects and province characteristics.

4.2   Sensitivity checks

To investigate the sensitivity of our main findings we performed a range of sensitivity analysis of which the results are presented in Table 5. For ease of comparison the first row of panel a reports the baseline OLS estimates while the first row of panel b reports the baseline 2SLS estimated, both presented in more detail in Table 4.

In panel a, we show the effects of omitting some of the provinces from our sample. Panel a2 shows the parameter estimates if we remove the province Friesland in which the language

spoken deviates most from standard Dutch. Neither for males nor for females the relevant parameter estimates are affected a lot. The effect of daily dialect-speaking for males is still significant negative while for females the effect is insignificantly different from zero. The provinces Noord-Holland, Flevoland, Utrecht and Zuid-Holland which together form the so called "Randstad" have dialects that do not differ a lot from standard Dutch. If we remove these provinces from the sample the effects hardly change either, as shown in panel a3.[8]

In panel b we show alternative 2SLS estimates. In panel b2, we show the relevant parameter estimates if we change the independent variable daily dialect-speaking to ever speaking-speaking, i.e. daily, regularly and once in a while. Daily usage of dialect can be self-selected to fit in a working environment, while ever usage of dialect during lifetime is probably more determined by location of origin. Ever dialect-speaking reduces male hourly wages with 9.7%. For females there is still no significant effect on hourly wages while the point estimates are once again very small.

In the fourth and fifth row of panel a we show the effect of dialect speaking on type of jobs. As with the OLS estimates, for both males and females, dialect-speakers on average end up with occupations at lower professional rank. And, daily dialect-speaking significantly lowers the probability of getting a non-manual job for males, but does not affect females' job type.

Panels b5 to b8 of Table 5 provide an indication of the heterogeneity of our main parameter estimates. Panel b5 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 more 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 a 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

---

[8]To investigate whether the wage penalty of dialect-speaking increases with the linguistic distance of the dialect we performed separate estimates for dialect-speaking workers. It turned out that the linguistic distance has no influence on the wage penalty.

longer significant negative effects on hourly wages. For females we find no effects irrespective of their age. Apparently, the wage penalty of dialect-speaking is more severe among younger males.

In panels b7 and b8, we distinguish workers according to their educational attainment. If speaking dialect is a potential signal of a lower productivity, we would expect that higher educated suffer more from dialect speaking because their targeted occupations require more intellectual skills. The magnitude of the wage penalty of dialect speaking is large for higher educated workers while for non-high educated workers the wage penalty disappears.

In an unreported sensitivity check, we control industry fixed effects of jobs, i.e. agriculture, industrial production, construction, retail trade, transportation, financial department, government services and so on. Regardless of industry fixed effects, the penalty of dialect-speaking on males' hourly wages is not substantially different. Besides, we can also control fixed effects for territorial divisions of provinces, that is, north, south, east and west. Geographic distance to Amsterdam is still a relevant instrument, and the effect of dialect-speaking is robust.

4.3   Interpretation

Having established that there is a significant wage penalty of dialect-speaking for males the question is how to interpret this finding. The penalty of dialect-speaking on monthly earnings does not work through employment probability or working hours, but dialect-speaking decreases hourly wages and the probability of getting a more profitable or non-manual job. This may be because these better paid jobs require proficiency in more widely used languages located in urbanized areas. It may be that dialect-speakers perform worse in terms of Dutch language proficiency. Yao et al. (2016) for example investigate the relationship between dialect speaking and academic performance of 5-6 year old children in the Netherlands. They find that dialect speaking has modest negative effects on language skills of young boys while young girls are not affected. However, it is not clear that this effect persists over time. Giesbers et al. (1988) reporting the results of an investigation at primary schools in two

dialect areas in the southern parts of the Netherlands conclude that dialect-speaking is not an educational disadvantage per se. However, they do find that dialect-speaking children are disadvantaged in terms of actual school performance and their choice of secondary education. One of the surprising findings is that school teachers seem to be prejudiced against dialect-speaking speaking children. Essays of dialect-speakers were graded worse than essays of non-dialect speakers if teachers were aware of the dialect-speaking of their pupils. However, if graders were unaware of the dialect-speaking they did not give lower grades to dialect-speaking children.

In our analysis, we focus on native Dutch who report having no problems in reading and speaking standard Dutch. So it is unlikely that dialect-speakers have a lower productivity at the workplace because of poor reading or writing skills. 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 employers 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. Therefore, mechanisms such as discrimination and signaling can explain the wage penalty better than actual productivity differences.

The fact that the wage penalty is only present for males but not for females may be related to gender difference in locational job choice. Analyzing LISS data, we find that females are more likely to work closer to home than males. The distance from work to home and commuting time for females are 10.5 km and 20.6 minutes on average, whereas for males this is 16.3 km and 27.5 minutes. Therefore, one would expect a smaller wage penalty of dialect-speaking for females since they are less likely to work in areas where their speech is considered to be a dialect.

5   Conclusions

We investigate whether daily dialect-speaking affects labor market performance of individual workers focusing on hourly wages. Using data from the Netherlands we conclude that dialect-speaking does not affect the probability that an individual has a job or conditional on having job weekly hours of work. However, male workers who speak a dialect daily on average have significantly lower hourly wages. For females we also find a negative effect of dialect-speaking on earnings but these effects are not significantly different from zero. Young and high educated male workers face a larger wage penalty of dialect speaking than their counterparts. Finally, conditional on other personal characteristics including educational attainment we find the daily dialect-speakers are less likely to have a high-ranked profession and more likely to work in manual jobs. To the extent that dialect-speaking also affects educational attainment our estimates are a lower bound of the true effect of dialect-speaking. Our data do not allow us to make a distinction between the various mechanisms that lead to dialect-speaking having negative wage effects for male workers. We can only speculate that prejudiced employers may be a reason. Or, self-prejudiced workers are responsible for the wage penalty of dialect-speaking to the extent that workers who speak a dialect self-select in worse jobs.

Our findings are all the more surprising since we use an imperfect indicator for dialect-speaking. Our identifying information is whether individuals speak dialect daily without reference to home or work. It could be that dialect-speakers do not speak dialect while at work. Furthermore, an individual who claims to speak standard Dutch may not speak a dialect according to someone from a different area in the Netherlands. Clearly, further research is needed in which more information is collected about speech patterns of standard Dutch speakers and dialect speakers preferably making a distinction between home and work and preferably not only relying on self-reported data.

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 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.

Cheshire, J., V. Edwards, H. Munstermann, and B. Weltens (1989). *Dialect and Education: Some European Perspectives*. Multilingual Matters Ltd.

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.

Driessen, G. (2005). In Dutch? Usage of Dutch regional languages and dialects. *Language, Culture and Curriculum 18*, 271–285.

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.

Giesbers, H., S. Kroon, and R. Liebrand (1988). Bidialectalism and primary school achievement in a Dutch dialect area. *Language and Education 2*, 77–93.

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.

Grogger, J. (2014). *Speech and Wages*. University of Chicago.

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

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory 10*, 707–710.

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.

Swadesh, S. (1952). Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society 96*, 121–137.

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. *Economics of Education Review 54*, 1–15.

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: Details on the variables

A1. Explanatory variables

Personal characteristics

- Dialect daily: dummy variable, 1 if the answer to the question "Do you ever speak dialect?" is "Yes, daily", 0 if "Yes, regularly", "Yes, sometimes" or "No, never".

- Age dummies: dummy variables for age categories 25 to 34, 35 to 44, 45 to 54 and 55+. Reference group is age below 25.

- Education dummies: dummy variables for educational categories primary school, intermediate and higher secondary school, intermediate and higher vocational school, university or higher. Reference group is primary school education.

- Number of children: number of dependent children in the household.

- Partner: dummy variable for having a partner living at the same address.

- Urbanization: dummy variables for degree of urbanization based on inhabitants per square-kilometer in the municipality of residence (in parentheses); extremely urban (2500 or more), very urban (1500 to 2500), moderately urban (1000 to 1500), slightly urban (500 to 1000), not urban (less than 500). Reference group is extremely urban.

- Religious: dummy variable for having a religion.

A2. Province characteristics

- GDP per capita: GDP, value added at market prices of the total economy, per capita.

- Employment: number of residents having a paid work for at least one hour a week, including self-employees.

- Population: total number of people residing in a province by January 1.

- Area Main Roads: total area of main roads in km^2.

- Distance to Amsterdam: straight distance from a municipality to Amsterdam in 5 km.

Dependent variables

- Employment: dummy variable for having any type of paid work, including family business and self-employment.

- Monthly earnings: personal monthly gross earnings from employment.

- Working hours: average hour of work per week.

- Hourly wage: calculated based on monthly earnings and working hours.

- Professional ranking: based on 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

- Manual jobs: unskilled and trained manual work, semi-skilled manual work, agrarian profession and skilled and supervisory manual work. Non-manual jobs include all other jobs, i.e. academic, independent, supervisory, commercial or mental work professions at any level.

- Happiness: based on the question "On the whole, how happy would you say you are?" Measured on an ordinal scale from zero (very unhappy) to ten (very happy).

Appendix B: Determinants of net migration inflow

As indicated in the main text, we assume that people do not move away from the place of origin or if they do, they stay in their province or do this not for economic reasons. To investigate whether this assumption is plausible, Table 6 presents the correlation between internal migration and labor market factors at province level available from 2010 to 2013. The data is from Statistics Netherlands. The dependent variable is the net inflow rate of migration due to inter-municipality moves. 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.

Table 6: Determinants of net migration inflow

	Net inflow	
GDP per capita	-0.000	(0.000)
log(Total compensation)	30.214	(31.864)
log(Number of employees)	-22.643	(90.213)
log(Total working hours)	-36.693	(95.840)
log(Population)	-18.253	(42.030)
Year fixed effects	Yes	
Province fixed effects	Yes	
F-statistics for wage factors	0.55	
# Obs.	48	