IZA DP No. 7019

English Deficiency and the Native-Immigrant Wage Gap

Alfonso Miranda
Yu Zhu

November 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

English Deficiency and the
Native-Immigrant Wage Gap

Alfonso Miranda
CIDE and IZA

Yu Zhu
University of Kent

ABSTRACT

English Deficiency and the Native-Immigrant Wage Gap[*]

We focus on the effect of English deficiency on the native-immigrant wage gap for male employees in the UK using the first wave of the UK Household Longitudinal Survey. We show that the wage gap is robust to controls for age, region of residence, educational attainment and ethnicity. However, English as Additional Language (EAL) is capable of explaining virtually all the remaining wage gap between natives and immigrants. Using the interaction of language of country of birth and age-at-arrival as instrument, we find strong evidence of a causal effect of EAL on the native-immigrant wage gap.

Corresponding author:

Yu Zhu
School of Economics
University of Kent
Canterbury, CT2 7NP
United Kingdom
E-mail: yz5@kent.ac.uk

1   Introduction

This paper investigates to what extent English deficiency explains the native-immigrant wage gap for males in the UK, after controlling for age, region of residence, educational attainment and ethnicity. Previous work have found that, upon immigration, good command of the language of the destination county carries a positive return of around 15-25% for migrants in the labour market (see e.g. Chiswick 1991, Chiswick and Miller 1999, Dustmann 1994). In the UK Dustmann and Fabbri (2003) report a return of about 18-20% using the Family and Working Lives Survey (FWLS) as well as the Fourth National Survey on Ethnic Minorities (FNSEM). However, after controlling for the potential endogeneity of English proficiency they find no statistically significant effect. This may be due to the small sample size the wage equations were fitted on, with 250 observations for the FWLS and 920 observations for the FNSEM.

Our contribution to the literature is threefold. The Dustmann and Fabbri (2003) study analysed data collected in the first half of the 1990s. Since that time, the UK has received a significant inflow of new migrants from Eastern Europe, following the expansion of the European Union. Hence, it is interesting to revisit the topic and investigate whether the returns to English proficiency remain at the same level as those reported nearly 20 years ago. Second, our analyses are based on a larger sample of immigrants, enabling us to estimate tighter confidence intervals. Finally, we use an IV strategy to address the issue of potential endogeneity of our English language proficiency indicator. Endogeneity may arise from two different sources: (i) self-selection into treatment, and (ii) measurement error. We address the 'aggregate' endogeneity problem using an IV strategy, without making an attempt to disentangle the contribution of the two sources of bias. Following Bleakley and Chin (2004, 2010) and van Ours and Veenman (2006), we use language of the origin country as well as its interaction with age-at-arrival as instrumental variables for the subpopulation of immigrants. Our identification strategy effectively compares older and younger arrivers from non-English-speaking countries, after controlling for an age-at-arrival effect which is common for all immigrants regardless of their native language. This identification strategy delivers a Local Average Treatment Effect (LATE) that is straightforward to interpret for the subpopulation of immigrants affected by the

2

instrument. Hence, this is an improvement to alternative identification strategies that have been used in the past.

2   Data and set-up of the analysis

Our empirical analysis is based on the first wave of the UK Household Longitudinal Survey, a longitudinal survey of just over 30,000 households in the UK in 2009-2011, including around 4,000 from the ethnic minority boost sample. We focus on the native-immigrant wage gap of male employees aged 19-60. Natives are defined as UK-born whites who speak English as first language. Conversely, immigrants are defined as people who were born abroad to two non-native parents.[1] After excluding missing values and outliers of log real hourly wage[2], we end up with a sample of 6,959 males, of which 1,203 are immigrants.

Table 1 reports summary statistics by immigrant status. The raw native-immigrant wage gap is a highly statistically significant 0.159 log points in favour of natives. However, this does not take into account any composition effect. Indeed, these two samples have very different characteristics. For instance, male immigrants in the UK are on average younger, hold higher qualifications, and live disproportionately in London compared to white natives. 71% of the immigrants declare speaking English as Additional Language (EAL), while 87% of them were born in developing countries. A significant minority (18.5%) of male immigrants are white, of which 71% were born in European countries.[3]

3   Results and discussions

In a wage equation, we measure the immigrant wage effect by a dummy variable, with a negative coefficient indicating a regression-adjusted native-immigrant wage gap in favour of natives. In Table 2 we successively introduce control variables. The native-immigrant wage gap increases by 0.10 log points when differences in age profiles and region of residence are accounted for in column 1. Moreover, once highest qualifications are controlled for in column 2 the gap widens by another 0.02 log points. When ethnicity

---

1   Blackaby et al. (2005) have studied the employment and earnings differentials between whites and UK-born ethnic minorities using a decomposition approach. We only include non UK-born immigrants in the treatment group, in order to exploit the variation in English deficiency induced by the variation in the age-at-arrival of immigrants from non-English-speaking versus English-speaking countries.

2   We exclude the top and bottom 1% of wages by highest qualification and ethnicity.

3   27% of immigrants arriving in the UK since 2004 are white, of which 61% are from Poland.

dummies are added in column 3, the gap is reduced to 0.117 log points, only one-quarter smaller than the raw differential. This implies that even white immigrants (the reference ethnicity category) suffer from large wage penalties compared to natives, holding age, region and education constant.

We then explore to what extent English deficiency explains this remaining wage gap in the next two columns. When EAL is added in column 4, the gap is reduced to 0.018 log points in favour of immigrants and becomes statistically insignificant. This implies that the entire remaining wage gap is explained by English deficiency. When we further include a dummy for immigrating from a developing country[4] and dummies for age-at-arrival in the UK for immigrants (column 5), the immigrant coefficient even turns into a positive and statistically significant 0.132 log points while the EAL effect remains significant and is of the same magnitude as the raw wage gap.[5] The fact that all age-at-arrival dummies are strongly negative suggests that for white immigrants arriving in the UK before 10 (the omitted category), there is no disadvantage associated with being an immigrant, holding all else constant.

In Table 3 we address the potential endogeneity of EAL using both the full sample and the subsample of immigrants only. In column 1, we instrument EAL using born in a non-English-speaking country as well as its interaction with a dummy for age-at-arrival greater than 9. Figure A1 shows the regression-adjusted mean probability of EAL, with 95% confidence intervals, by age-at-arrival and home country language. Immigrants from non-English-speaking countries who arrived before the age of 10-14 are, statistically, as likely to be EAL as immigrants from English-speaking countries. In contrast, if immigration occurred after age 10-14 the two groups are statistically different. Following the theory of the critical period for second language acquisition Bleakley and Chin (2004, 2010) argue that, after controlling for educational attainment and other background variables, differences in English proficiency between immigrants from English-speaking and non-English-speaking countries before and after age 10 are uncorrelated with current wages because any non-language age-at-arrival wage effects are the same for all

---

4   The developing country dummy might partly capture the differences in quality of education.
5   We include age, age square and age-at-arrival in bands of 0-9, 10-15, 16-29 and 30+ to disentangle the effect of assimilation and effects of language. However, our identification only relies on the *interaction* between born in a non-English speaking country and age-at-arrival greater than 9.

immigrants regardless of their home country language. Notice that under the postulated identification strategy the IV estimator is analogous to a difference-in-differences estimator that calculates language wage effects net of age-at-arrival wage effects.

The bottom panel of Table 3 shows that the instruments are strong predictors of EAL status both individually and jointly. Being born in a non-English-speaking country increases the incidence of EAL by 33% while arriving in the UK after age 9 from a non-English-speaking country increases the probability by another 49%. Moreover, the model also easily passes the over-identification test. We find that EAL effect becomes larger when we allow it to be endogenous but remains statistically significant at the 5% level, although we won't be able to reject the null of equality with the OLS estimate. On the other hand, the immigrant wage effect remains virtually the same as in the OLS specification.

Column 2 shows that instrumenting using the interaction term only makes no difference. The just-identified IV is unlikely to be subject to a weak-instrument critique. GMM and LIML estimates (available upon request) also come out very similar, lending further support to the robustness of our IV results.

Columns 3 and 4 replicate the 2SLS estimation, using the subsample of immigrants only. The fact that two sets of results are virtually identical implies that the causal effect of EAL is identified by variation *within* the sub-population of immigrants in English deficiency induced by age-at-arrival between immigrants from English-speaking and non-English-speaking countries.

4   Conclusions

We find a composition-adjusted male native-immigrant wage gap in the UK of 12%, slightly below the raw wage differential. However, this gap virtually disappears after controlling for the EAL indicator. We address the endogeneity of EAL with an IV strategy and use born in non-English-speaking country and age-at-arrival as instruments for identifying a LATE that is straightforward to interpret for the subpopulation of first-generation immigrants affected by the instrument. Our IV regressions indicate that EAL

has a causal negative effect of -23% on wages, which is significant at 5%, and robust to various specifications.

References

Blackaby, D.H., Leslie, D.G., Murphy, P.D., O'Leary, N.C., 2005. Born in Britain: How are Native Ethnic Minorities Faring in the British Labour Market? Economics Letters 88, 370-375.

Bleakley, H., Chin, A., 2004. Language skills and earnings: evidence from childhood immigrants. Review of Economics and Statistics 86, 481 – 496.

Bleakley, H., Chin, A., 2010. Age-at-arrival, English proficiency, and social assimilation among US immigrants. American Economic Journal: Applied Economics 2, 165–192.

Chiswick, B.R., 1991. Speaking, reading, and earnings among low-skilled immigrants. Journal of Labor Economics 9, 149-70.

Chiswick, B.R., Miller, P.W., 1999. Language skills and earnings among legalised aliens. Journal of Population Economics 12, 63-89.

Dustmann, C., 1994. Speaking fluency, writing Fluency and earnings of migrants. Journal of Population Economics 7, 133-56.

Dustmann, C., Fabbri, F. 2003. Language proficiency and labour market performance in the UK. Economic Journal 113, 695-717.

Van Ours, J. C., Veenman, J., 2006. Age at immigration and educational attainment of young immigrants. Economics Letters 90, 310-316.

Table 1: Summary statistics, by immigrant status

	Natives (N=5756)	Immigrants (N=1203)	Native-immigrant gap
Log real hourly wage	2.504	2.345	0.159**
EAL	0	0.710	-0.710**
Born in developing country	0	0.874	-0.874**
No qualification	0.141	0.162	-0.022*
Below GCSE/O-Level	0.094	0.076	0.017**
GCSE/O-Level	0.286	0.081	0.206**
A-Level	0.119	0.097	0.022**
Higher Education Diploma	0.089	0.086	0.004
First Degree	0.173	0.212	-0.039**
Higher Degree	0.097	0.285	-0.187**
Highest qualification is foreign	0.002	0.392	-0.389**
Age	40.3	37.5	2.8**
White	1.000	0.185	0.815**
Mixed	0	0.013	-0.013**
Asian	0	0.552	-0.552**
Black	0	0.140	-0.140**
Other Ethnicity	0	0.111	-0.111**
London	0.061	0.485	-0.424**
Southeast	0.132	0.069	0.063**
Rest of England	0.615	0.391	0.223**
Wales	0.051	0.009	0.042**
Scotland	0.098	0.030	0.068**
Northern Ireland	0.044	0.016	0.028**

Note: **(*) = significant at 5% (10%) level based on Welch's t-test.

Table 2: OLS wage equations, pooled natives and immigrants sample (N=6959)

	(1)	(2)	(3)	(4)	(5)
Immigrant	-0.258 (13.7)**	-0.276 (14.0)**	-0.117 (3.2)**	0.018 (0.4)	0.132 (2.3)**
EAL				-0.198 (6.0)**	-0.158 (4.4)**
Born in developing country					-0.092 (1.8)*
Age-at-arrival 10-15					-0.136 (2.2)**
Age-at-arrival 16-29					-0.090 (2.0)**
Age-at-arrival 30+					-0.172 (3.2)**
Highest qualification dummies	no	Yes	yes	yes	yes
Ethnicity dummies	no	No	yes	yes	yes

Note: Absolute t-statistics based on robust standard errors in parentheses; **(*) = significant at 5% (10%) level. Other controls include age, age squared and region dummies.

Table 3: 2SLS wage equations

	Full sample		Immigrants only	
	IV (over-identified)	IV (just-identified)	IV (over-identified)	IV (just-identified)
Immigrant	0.151 (2.5)**	0.151 (2.5)**		
EAL	-0.253 (2.4)**	-0.251 (2.3)**	-0.232 (2.2)**	-0.242 (2.3)**
Born in developing country	-0.061 (1.0)	-0.062 (1.0)	-0.085 (1.4)	-0.081 (1.3)
Age-at-arrival 10-15	-0.110 (1.6)	-0.111 (1.6)	-0.145 (2.1)**	-0.142 (2.0)**
Age-at-arrival 16-29	-0.056 (1.0)	-0.057 (1.0)	-0.076 (1.3)	-0.072 (1.3)
Age-at-arrival 30+	-0.135 (2.0)**	-0.136 (2.1)**	-0.125 (1.8)*	-0.121 (1.8)*
p-value for over-identification test	0.887	-	0.612	-
First stage (dependent variable = EAL)				
Born in non-English-speaking country	0.330 (2.7)**		0.332 (2.8)**	
Born in non-English-speaking country * (age-at-arrival>9)	0.487 (4.1)**	0.764 (19.0)**	0.490 (4.3)**	0.767 (18.8)**
F-stat for exclusion restrictions (p-value)	248.9 (0.000)	359.7 (0.000)	228.0 (0.000)	353.9 (0.000)

Note: Absolute t-statistics based on robust standard errors in parentheses; **(*) = significant at 5% (10%) level. Other controls include age, age squared, region dummies, highest qualification dummies and ethnicity dummies.

APPENDIX

Figure A1: Fitted EAL probability by age-at-arrival and home country language