Economic Analysis and Policy 52 (2016) 66–78

Contents lists available at ScienceDirect

Economic Analysis and Policy

journal homepage: www.elsevier.com/locate/eap

Full length article

Effects of language proficiency on labour, social and health outcomes of immigrants in Australia

King Wa Tam [a], Lionel Page [b,*]

[a] University of Queensland, School of Public Health, Herston Qld, 4006, Australia

[b] Queensland University of Technology, School of Economics and Finance, 2 George Street, 4000, Brisbane, Queensland, Australia

ARTICLE INFO

Article history:
Received 4 May 2016
Received in revised form 22 August 2016
Accepted 22 August 2016
Available online 1 September 2016

Keywords:
English proficiency
Childhood immigrants

ABSTRACT

We investigate the causal effect of English proficiency on labour, social and health outcomes of immigrants in Australia. We use age at arrival combined with country of origin to form an instrument of English proficiency. We find that immigrants in Australia with better language proficiency are able to earn higher income, attain higher level of education, have higher probability of complete tertiary studies, and get more hours of work per week. Language proficiency also improves social integration, leading to higher probability of marriage to a native and higher probability of obtaining citizenship. We find only limited evidence with respect to the hypothesised causal relationship between language and health for immigrants. This last result may be due to small sample sizes.

1. Introduction

A large proportion of the Australian population are foreign-born: the latest estimate from the Australian Bureau of Statistics (ABS) indicates that 27%[1] were born overseas, one of the highest percentages in the world (ABS, 2011). This figure is higher than many other major countries that also have had a long history of immigration such as Canada, the United States and New Zealand.[2]

The objectives of this study are to investigate the extent to which language ability influences immigrants' lives in a wide range of domains: labour market outcomes (e.g. income, employment, hours of work and education), social integration (e.g. marriage to a native, citizenship, and voluntary work), and health conditions (e.g. degree of health care access, and healthy lifestyle choices).

Early studies have used the ordinary least squares (OLS) methodological approach to analysing data on migrants. For example, Mcmanus et al. (1983), Kossoudji (1988), Tainer (1988), Rivera-Batiz (1990), Chiswick (1991) and Dustmann (1994), have analysed how language ability, or the lack thereof, impacts employment status and income of immigrants in countries such as the United States and Germany. They generally found that the lack of host-country dominant language skills would constrain wage rates of those who are foreign-born, and that the wage gap between the natives and foreign-born narrows as immigrants attain more language skills.

* Corresponding author.
E-mail address: lionel.page@qut.edu.au (L. Page).

1 This figure is valid as of 30, June 2010.

2 21.3% for Canada, 22.4% for New Zealand, 13% for the United States.

http://dx.doi.org/10.1016/j.eap.2016.08.003
0313-5926/

As researchers began to take note of the potential biases produced by treating language skills as an exogenous variable in their estimation of the effects, many employed an instrumental variable approach. Some examples are Chiswick and Miller (1995), Angrist and Lavy (1997), Dustmann and Van Soest (2002), Shields and Price (2002), Dustmann and Fabbri (2003) and Bleakley and Chin (2004, 2010). The bias has been determined to be mainly due to measurement error, but there is also a possibility of unobserved non-language ability and reverse causality. These studies using instrument variable (IV) strategies have found a significant positive impact of proficiency on labour outcomes, but they also point out the bias of OLS estimates due to endogeneity.

In the present study we implement the empirical strategy proposed by Bleakley and Chin (2004, 2010). They used the fact that the age of arrival of migrants has a critical impact on their ability to learn the new language in order to identify the effect of English proficiency on market outcomes in the USA. The difficulty to learn a new language at a later age can be seen as providing a source of exogenous variation in language proficiency.

We use this same technique to study the causal effect of English proficiency on a range of outcomes of Australian immigrants. As in previous studies on migrants' proficiency we investigate labour market outcomes. We complement this study with an investigation of health outcomes on a smaller dataset. Whilst the evidence is less clear on the health data, to our knowledge it is the first such study and which therefore gives it some significance. To proceed, we used two datasets. For the labour market outcomes and social assimilation, data from the Australian 2006 Census are used, whilst data from the Australian National Health Surveys in 2001, 2004–5 and 2007–8 are used for the study on health outcomes. The demographic composition of the estimation samples consists of childhood immigrants—adult immigrants who arrived in Australia during their childhood along with their parents. This should not be confused with child immigrants who are currently still young.

The remainder of the paper is as follows. Section 2 presents the methodology, Section 3 describes the data used, Section 4 presents the study of the effect of proficiency on labour market and social outcomes, Section 5 looks at effects on health outcomes and Section 6 concludes.

2. Methodology

2.1. The issue of endogeneity

Following the recent works of Bleakley and Chin (2004, 2010), we have a model based upon a linear regression framework of the labour market outcomes y_{ija} of an individual i, who was born in the country j and who arrived in Australia at age a:

$$y_{ija} = \alpha + \beta \mathrm{ENG}_{ija} + \delta_a + \gamma_j + \rho w_{ija} + \varepsilon_{ija} \tag{1}$$

where ENG_{ija} is a measure of English language skills, δ_a is a set of age-at-arrival dummies, γ_j is a set of country-of-birth dummies, and w_{ija} is vector of individual characteristics (i.e. age, age squared, and gender) and survey year dummies where applicable.[3] The coefficient β is the parameter of primary interest and represents the impact of English language skills on labour market success. However, obtaining consistent estimates of β is difficult because ENG_{ija} is likely to be endogenous and is likely to contain measurement error. Endogeneity and measurement error render ordinary least squares (OLS) estimates both biased and inconsistent, so any inference based upon those results is unreliable.

One potential source of endogeneity is omitted variables. For example, an individual's unobserved ability (e.g. intelligence, aptitude, etc.) is likely to positively impact labour market outcomes, and also language proficiency. This correlation pattern would result in an upward bias in the least squares coefficient estimate for English proficiency. Another potential source of endogeneity could be reverse causality. That would occur if our outcomes of interest directly impact upon ENG_{ija}. Suppose y_{ija} is the amount of work in hours, it is possible that an individual acquires further language improvement due to close proximity with native speakers in a work environment.

Measurement error is also possible because the measure of language skills is likely to be composed of the unobserved true value and some noise. When estimating the gradient for the OLS regression, the effect of English skills will be diluted by the very existence of this noise to the extent of its variability. This is also known as attenuation bias. English ability is arguably difficult to measure precisely even when comprehensive tests are employed, so it would be expected that a simple self-reported answer in surveys or interviews contains substantial measurement (Dustmann and Van Soest, 2002).

Since the omitted variables bias is upward and the attenuation bias is downward, the direction of the overall bias in the OLS estimate is difficult to ascertain a priori. However, Bleakley and Chin stated in their 2004 paper that the upward endogeneity bias in language skills is outweighed by the enormous downward attenuation bias. Using US national literacy test scores as the true value for English language skills, they verified the measurement error was indeed mostly classical, and that the IV estimate for the effect of English ability on earnings was higher than the OLS estimates. Likewise, Shields and Wheatley (2002) as well as Dustmann and Fabbri (2003) found downward bias in their OLS estimates. Similarly, given these results we would expect our OLS estimate is biased towards zero.

[3] Survey year dummies are included for estimation that uses the National Health Surveys.