Intergenerational assimilation of UK immigrants in the labour market: A minor assumption with enormous implications for inference☆

Zovanga L. Kone

Centre on Migration, Policy and Society (COMPAS), 58 Banbury Road, University of Oxford, OX2 6QS, United Kingdom

HIGHLIGHTS

- The economic integration of UK's immigrants and their children is a key policy issue.
- How researchers define nativity status can affect the inference of this integration.
- It can cause classification errors in samples, which might sternly bias estimates.
- We present analytical results showing these biases can go in any direction, even in simple models.
- The empirical analysis provides supportive evidence for the analytical results.

ARTICLE INFO

ABSTRACT

Studies on the intergenerational assimilation of UK immigrants and their UK-born children have mainly relied on ethnicity and birthplace to measure nativity status because of data limitations. This would inevitably lead to classification errors in the sample. We present analytical results showing biases resulting from classification errors can go in any direction even when the sole regressor is a binary variable. The empirical analysis confirms such unpredictable implications for inference. A more accurate measure of nativity status based on parent's birthplace indicates the integration of immigrants might be different to what we would get from a measure prone to wrongly classifying individuals.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

"The issue of immigration and the integration of immigrants and their children are high on the policy agenda of the EU and OECD countries, both from an economic and a social standpoint. The active participation of immigrants and their children in the labour market and, more generally, in the public life is vital for ensuring social cohesion in the host country..." [OECD (2015). "Indicators of Immigrants Integration, Settling In" p. 9.]

The last decade has been marked by an increased interest in the economic integration of the children born to immigrants in the UK, not least because they and the foreign-born population represent over 15% of the country's working age population. The focus has often been from an intergenerational perspective (see Algan et al., 2010; Dustmann and Theodoropoulos, 2010; Heath et al., 2008). A key problem that arises in the empirical analysis of the presence (or absence thereof) of intergenerational integration of UK immigrants in the labour market, however, is that the survey widely used to examine such question, i.e. the UK Labour Force Survey (LFS), lacks information on parental country of birth. To circumvent this data limitation, a common assumption in the literature has been to categorise UK-born individuals of white ethnicity as natives and those of non-white ethnicity as second-generation immigrants.[1] Such assumption would inevitably lead to classification errors in the sample: the UK Household Longitudinal Study (HLS), an alternative survey to the LFS, suggests over 10% of non-white UK-born have UK-born parents, hence by definition not second-generation immigrants.

[1] An exception is Platt (2005), who uses information on parental birthplace when examining occupational class mobility amongst UK-born individuals of Indian and Black-Caribbean origins.

Although this has been acknowledged in the literature (Heath and Li, 2008), the implications for inference have not been fully considered. As Kreider notes (2010), "…measurement error in a binary regressor can result in severe identification deterioration of the regression coefficients in the presence of very few classification errors…". Klepper (1988) show coefficient estimates on dichotomous variables measured with error might no longer provide lower bounds for true parameters in multivariate regressions, hence the bias could be non-attenuating.

We first present analytical results showing that even in regressions where the sole regressor is a binary variable, biases due to classification errors can go in any direction. The empirical analysis shows how small amounts of classification errors in samples, which arise from nativity status measures based on ethnicity and birthplace, would lead to large statistical biases in estimates of interest — we look at employment probabilities as outcome. When we measure nativity using information on parental birthplace and compare second-generation immigrants to their native counterparts of the same ethnicity, two groups susceptible to misclassification, we find the former may have better labour market outcomes than the latter, a new insight.

2. The data

We use data from the LFS and the HLS. The LFS is conducted by the UK's Office for National Statistics and primarily collects information on economic activity. It is the country's largest household quarterly survey and follows respondents over five quarters. The HLS, on the other hand, is the largest longitudinal household survey of its kind and covers broader socio-demographic variables. It is conducted annually, with the Universities of Essex, Warwick and the London School of Economics providing scientific leadership for the survey. Importantly, unlike the LFS, it contains information on parental birthplace, thereby permitting a more appropriate identification of natives and second-generation immigrants.

Immigration related studies mostly use the LFS because it offers larger samples of immigrants and ethnic minorities: whereas the HLS started in 2009, with the main sample comprising 49,920 surveyed addresses (about 40,000 households), the LFS began in 1973 and a quarterly sample presently contains about 40,000 households. Both surveys, nevertheless, aim to be fully representative of the UK's population[2] (Knies, 2016; ONS, 2001) and population estimates from both are comparable, with differences generally small wherever these exist (Buck and McFall, 2011; Burton et al., 2011) — see appendix for comparison Table A1. The HLS therefore makes it possible to gauge the degree of misclassification if using LFS samples.

We pool LFS data from 1993 to 2007 (as in Algan et al., 2010) and HLS data from 2009 to 2014. The samples are restricted to 16–64 year-olds. Following the literature, we first define (1) a foreign-born as first-generation immigrant; (2) a UK-born of non-white ethnicity as second-generation immigrant; and (3) a UK-born of white ethnicity as native. Later, when using HLS data, we distinguish second-generation immigrants from natives based on parent's birth country. Sample summary from the LFS is reported in Table 1 and Table 2 reports nativity status distributions of UK-born individuals by ethnicity, using HLS data. Although all groups are affected, the extent of classification errors is particularly startling among the Black-Caribbean ethnicity group: about a quarter have UK-born parents (see last two columns of Table 2).

Table 1
Summary statistics of sample, LFS (1993–2007).

Generation status	Proportion (%) Both gender		Employment probabilities			
			Male		Female	
	1st	2nd	1st	2nd	1st	2nd
Natives	90.3		78.3		78.3	
Immigrants	8.4	1.8	73.1	59.5	55.2	53.1
of which						
White	57.5	–	77.1		63.6	
Indian	14.9	32.1	78.8	63.1	54.9	57.7
Pakistani	8.5	21.9	63.8	49.8	16.4	35.1
Black-African	7.1	7.6	61.0	60.2	47.6	55.7
Black-Caribbean	5.0	30.8	64.6	65.2	60.8	61.6
Bangladeshi	3.5	4.1	55.8	41.3	12.8	35.2
Chinese	3.5	3.6	63.1	58.6	52.1	60.6

Note: Sample designed to replicate Algan et al. (2010). Nativity status measured using ethnicity and birth in the UK. Sample comprises 16–64 year olds.

3. Empirical analyses and results

We draw on Algan et al. (2010) [abbreviated as ADGM hereafter] as a benchmark in terms of inferential method with LFS data and the ethnicity groups looked at: Bangladeshi; Black-African; Black-Caribbean; Chinese; Indian; Pakistani; and White.[3] Intergenerational integration (or lack thereof) is inferred using estimates from two rounds of models: (1) a comparison of first-generation immigrants from different ethnic groups with natives; and (2) a comparison of second-generation immigrants from different ethnic groups with natives. Estimates from (1) and (2) for the same ethnic group are then compared to infer intergenerational integration.[4] The estimated equations are as follows:

$$Y = \alpha + \beta X + \theta_e D_e + \varepsilon \tag{1}$$

Y denotes employment status. Individuals in employment (including unpaid family workers) are given the value of 1 and the remaining value 0 — reason for economic inactivity is not considered. Those in full-time education are excluded from the estimation. X comprises control variables (i.e. gender, educational attainment levels, region of residence, year dummies and potential experience). D_e is a vector of dummies, which take value 1 for a non-native of a given ethnicity group. ε is the error term. How sensitive estimates of θ_e are to misclassifications in the sample is the object of interest.

We first present analytically the implications of classification errors in the sample for inference, followed by empirical evidence.

3.1. Implication of classification errors

Consider a simple linear probability model with one binary regressor:

$$y = \alpha + \beta d^* + \varepsilon \tag{2}$$

where d^* is a dummy, which is 1 if the individual is a second-generation immigrant. However, the econometrician only observes a mismeasured value of d^*, say d, with error (v) such that: $d =$

---

[2] Addresses surveyed for the HLS are drawn from a stratified sample of postcode sectors, which are at greater geographical aggregations level than those used for the LFS.

[3] Those of mixed race, other Asian/Black ethnicities are excluded as in ADGM.

[4] Intergenerational mobility in labour market outcomes is generally examined using two main approaches (Gregg et al., 2017). The first approach compares outcomes of a parents' generation and those of their children generation. The second examines how childhood circumstances affect an individual's outcomes in their adult life. The main approach adopted in the literature on UK immigrants is the first one.