Economic Assimilation of Pre- and Post-2004 EU Enlargement Immigrants to the UK

Giuseppe Forte
University of Warwick

Abstract

I employ the first five waves of the UK Household Longitudinal Study to test the Immigrant Assimilation Hypothesis – i.e. whether immigrants converge to native labour market achievement. In addition, I test whether assimilation is different for post-2004 immigrants. For earnings, I find an average yearly assimilation rate in excess of 2%, which bridges the 32% earnings deficit at entry in around 25 years – although dividing immigrants by region of origin shows a great amount of heterogeneity. Immigrants do not present lower probability of employment at entry, with some groups instead being more likely to find employment than natives.

Date: September, 11^{th} 2016. Student Number: U1551895.
I would first like to thank Jonathan Portes, whose guidance as supervisor was essential for the completion of this dissertation. My gratitude also goes to Marcus Miller – who kindly interceded to Jonathan making external supervision possible – and Divya Deepthi, for being a patient interlocutor.

Contents

List of Figures

List of Tables

1 Introduction

The number of migrants residing in the United Kingdom rose from 4.4 million (7.6% of the population) in 2001 to 8.3 million (13%) in 2014, the increase being roughly evenly split between arrivals from the European Union and the rest of the world. While a lot has been written on the labour market (Dustmann, Frattini, Preston (2013); Manacorda, Manning, Wadsworth (2012)) and fiscal (Dustmann and Frattini (2014)) impact of such sharp rise in immigration, the economic fortunes of foreign-born residents have attracted scarce attention from the British academia.

Such lack of interest is idiosyncratic to the United Kingdom, as large literatures exist in countries with similar histories of migration inflows; more importantly, it is unmotivated: labour market integration of immigrants is in fact closely related to their net contribution to the country of destination, and as such it should be – especially in a post-Brexit scenario that entails a reform of immigration policy – at the centre of policymakers' and public attention. The debate on migration that took place in the months leading to the June, 23^{rd} vote, largely ideological and mostly bereft of agreed notions[1], serves as a reminder of the necessity for more research on recent labour flows. It is telling that Chiswick (1980), Bell (1997) and Wheatley Price (2001) are the only UK-focused studies that investigate the existence and extent of economic assimilation, i.e. the convergence of immigrants to native levels of labour market achievement (in terms of earnings or employment rate) as years of residence accumulate, published before the May 2004 European Union enlargement[2]. Such event and the migratory

[1]When Lord Beecham submitted a Parliamentary Written Question (HL5912) on February, 4th 2016 regarding "annual benefits paid to EU migrants in the UK" and their "contribution [. . .] to the public purse through income tax recepits and VAT" Lord O'Neill – current Commercial Secretary to the Treasury – replied "The information is not available" (Beecham (2016a)). No estimate of VAT paid by EU nationals was available by June, when the question was asked again (Beecham (2016b)).

[2]I do not report a summary of the political debate around the Enlargement and the British position with regards to it for brevity reasons. Please refer to the "Review of the Balance of Competences between the United Kingdom and the European Union - EU Enlargement" report issued by HM Government in December 2014 for more information and a brief history of enlargement.

inflows that followed, coupled with the increased availability of longitudinal administrative data, renewed (albeit still minor) interest in the experience of immigrants, principally embodied by Dickens and McKnight (2008) and Lemos (2013).

The consensus emerging from this restricted literature is that immigrants do in fact converge to native levels of earnings and employment, even though with significant heterogeneity between origin groups, arrival cohort and genders. However, estimates vary significantly, the fundamental reason being the different datasets used due to the lack of a clearly superior source: the first three mentioned studies employ either a single or multiple cross-sections of survey/census data (1970 Census, General Household Survey and Labour Force Survey respectively), which provide more detailed personal information from a sample that is randomly selected – and thus different – in each wave; the latter two make use of the Lifetime Labour Market Database, which is both long and large but does not include fundamental information such as education.

In this paper I make use of an alternative and so far under-explored dataset in investigating the assimilation hypothesis, the Household Longitudinal Study (*Understanding Society*). I restrict my analysis to individuals aged 18-65 who have been interviewed for all of the five waves published: this selection yields a sample of 14188 adults, 2344 of which foreign-born – for a total of 70940 individual-wave observations. By analysing these data, I test whether immigrant earnings converge to native earnings over time and, similarly, whether there is convergence in the probability of employment. What distinguishes this study from the previous UK literature is that for the first time I am able to analyse longitudinally a sample of 514 immigrants who arrived after the 2004 EU enlargement, whereas datasets used beforehand – that only extended until 2006 – did not allow to include newcomers' profiles. This allows me to shed some preliminary (as studies with a larger representative sample are sorely needed) light on the assimilation paths of post-2004 immigrants.

In accord with past British studies, I find robust evidence of positive earnings assimilation – at a rate of over 2% per year – and an income gap at entry of over 30%, with large differences by region of origin. Following these estimates, it takes around 25-26 years for an average immigrant to obtain the wage of a native with similar specifics. After dividing potential experience in experience acquired in the country of origin and in the UK, I find insignificant returns to the former and no evidence of the relationship between the two found by Bell(1997). Workers who arrived to the UK after 2004 enter the market at an insignificant further deficit when compared to previous immigrants, and their rate of assimilation is not found to be different than that of older immigrants. No such entry gap and rate of convergence are found analysing the probability of employment: some origin groups are even found to be more likely than natives to find a job as they move into the British labour market, with pre-migration experience increasing their prospects more than British experience.

The dissertation is structured as follows: Section 2 reviews some international studies on assimilation, introducing some fundamental terminology and concepts, to then present results from the British literature and past longitudinal studies; Section 3 presents the dataset and some descriptive statistics, commenting on modifications and variables derived from information in *Understanding Society*; Section 4 presents the methodology used for earnings and employment assimilation analysis; Section 5 presents estimation results and analyses their meaning and implications; Section 6 summarises the results and concludes the dissertation. *In fine*, the Appendices gather more tables, robustness checks, additional sample descriptives and information on the variables used.

2 Literature Review

2.1 Introduction to the debate

The first stone of the debate on the economic assimilation of immigrants was laid in Chiswick (1978) where, analysing US data, the author defines the process in question as a result of the *Americanization* of the newcomer. Based on the Human Capital Theory framework developed in Becker (1964), Chiswick argues that immigrants enter the labour market at a disadvantage, as the returns to notions they bring from their native country are discounted due to partial incompatibility, and they thus need to invest in capital specific to the new market to improve their condition. Such investment can depress initial earnings by not allowing the individual to dedicate the whole of their time to work, but it improves future prospects - especially as it tends to be less firm-specific than knowledge acquired by natives in the first years of more stable employment.

In addition, the author imagines migration costs as a fixed proportion of native-country wages, which will imply positively selected immigration (i.e. higher average human capital among migrants than the population average in the country of origin) as more able and motivated workers experience returns to ability in the new market that increase at a higher rate than the cost of migrating. The combined effect of initial investment and positive selection is a steeper-than-average experience-earnings schedule among immigrants, who are expected to start with lower wages to then overtake natives as years since arrival (YSA) increase.

Using a cross-sectional subsample of the 1970 US Census of Population containing 34321 native and 1924 foreign-born men, Chiswick finds that being born abroad is associated with a significant negative wage gap at entry that is then reduced along time, at a slightly declining rate of around 1.5% per year. Most importantly, he argues that the gap at entry is not only eliminated, but also reversed as established immigrants achieve higher salaries than natives after 10-15 years, which can seemingly support the ability-motivation hypothesis of immigrant positive selection.

While largely influential in the way assimilation is framed and reconciled with theory, the paper has some clear limitations, and the interpretation of the results was soon disputed by many. The points of dispute are best summarised by Borjas (1985). First, being a cross-sectional study, Chiswick's paper could not account for return migration, which constitutes a further level of selection than the initial migration decision: if immigrants return to their native country (they *remigrate*) in a non-random way and remigrations increase along time for each wave, the cross-section analysed will not be representative of the immigrant population as a whole but only of self-selected 'survivors'. Depending on whether the selection is positive or negative, estimates of returns to YSA will be biased upwards or downwards. Since data on emigration in the 1980s were largely latent, the extent of remigration bias was not fully gauged until the analysis of administrative longitudinal data made in Lubotsky (2007).

Instead, the main contribution against the straightforward interpretation of Chiswick's results is made by introducing the concept of *cohort quality*: the observed convergence, Borjas argues, is not true assimilation but a result of declining human capital levels among more recent migration waves, levels that are instead implicitly assumed not to vary significantly in Chiswick's research.

To tell the effects apart, Borjas resorts to using two (and more in later studies, cf. Borjas (1994); Borjas (2013)) waves of US Census, which allows him to follow the decennial growth in earnings for synthetic cohorts (i.e. cohorts in which the observed unit is not the individual, as censuses have no cross-wave identifier, but the population of those who were born in a set period of time). He can then distinguish within-and across-cohort earnings growth: comparing the former to native earnings dynamics serves to quantify assimilation with coeval Americans; the latter instead measures the difference in market returns to cohort-specific attributes (which could be due to market changes – e.g. a fall in demand for immigrant labour – or differences in 'quality', even though Borjas seems more convinced by the second).

Estimates of within- and across-cohort earnings growth lead Borjas to conclude that there has been a secular decay in the quality of immigrants to the US, reconciled with the 1965 Immigration and Nationality Act shift in immigration policy – which caused most of permissions to be granted for family reunification, superseding previous more labour-market-oriented policies – and possible increases in illegal migration. He finds cross-sectional studies *à la* Chiswick overestimate absolute and relative (when compared to natives with similar characteristics) earnings growth of migrants, with some cohorts not experiencing convergence to native earnings levels. The latter finding is corroborated in Borjas (2013) by using more Census waves: the study in addition links slower assimilation (or complete lack thereof) to slower human capital growth (proxied by cohort growth in English proficiency) and belonging to a larger national origin group, which could allow immigrants to work and create networks in sub-economies with different characteristics from that of the natives.

2.2 Early British evidence

While concerns over the declining quality of immigrants to the United States – nowadays widely acknowledged – spurred research (cf. Funkhouser and Trejo (1995), LaLonde and Topel (1992) for studies that reach more nuanced conclusions than Borjas(1985)), comparable studies for the United Kingdom have been relatively scarce – even more so before the 2004 enlargement of the European Union.

Chiswick (1980) first tried to replicate the study conducted in 1978 by using data from the 1972 UK General Household Survey (GHS): although results are in line with his previous paper (immigrants start at a significant earnings disadvantage and catch up over time), the paper suffers from a very small sample of immigrants (only 341) in addition to the intrinsic problems of cross-sectional studies. In the following years, the focus of most British studies shifted to ethnic differences in earnings and employment (cf. Stewart (1983); Blackaby (1986); Blackaby te al. (1994)), and immigrant economic performance was left unexplored until following decade.

Bell (1997) was the first to exploit the wealth of cross-sectional data provided by multiple waves of the General Household Survey: he pools 20 consecutive surveys (1973-1992) to explore the origin mix, educational attainments and wages of immigrants. In his analysis of wages, which owes to Borjas (1985) in distinguishing cohort and assimilation effects, Bell finds very different gaps at entry and assimilation rates among ethnic groups: while experienced Black immigrants start with lower wages that anyway rise faster than those of native workers and inexperienced Black immigrants enter with no wage differential, White immigrants are found to enter the market at an initial wage premium (possibly as a result of positive selection) eroded over time.

This suggested inability to reconcile textbook expectations of economic assimilation with recent UK migrant experience has been confirmed by Clark and Lindley (2009), who use 1993-2004 Labour Force Survey (LFS) data: in addition to confirming the persistent wage differences between Whites and Non-Whites in the British market, the authors find "considerable diversity in the patterns of immigrant earnings and employment assimilation" (pg. 196) among ethnicities.

2.3 Longitudinal data in Assimilation Studies

A parallel strain of the assimilation literature has recently tried to measure economic convergence by using longitudinal data; however, chiefly due to the costs associated with collection, such datasets have historically been for many countries scarce and often comprising less individuals than cross-sectional sets. Panels are particularly apt for these studies as their analysis can obviate some of the issues associated with cross-sectional and pseudo-panel (i.e. iterated cross-sections) data: first, the use of the Within-Group estimator can net out cohort-specific, year-of-entry and more generally time-invariant effects; secondly, cross-wave identification allows, in particularly rich datasets, to account for re-migration and sample attrition.

Borjas (1989), using the 1972-1978 waves form the US Survey of Natural and Social Scientists and Engineers, looks to estimate assimilation among 15362 professionals (1166 of which foreigners), confirming his previous finding that more recent cohorts do not converge to native earnings. In addition, assuming that the rate of return migration can be grossly captured by the sample attrition differential among foreign-born and natives, the author concludes that re-migration is negatively selected, as immigrants who leave the sample tend to have lower wages in 1972.

For Germany, Pischke (1992) analyses the economic assimilation of *Gastarbeiters* using 1980s data from the Socio-Economic Panel. Using a sample of 2976 workers, 858 of which foreign-born, he finds no sign of assimilation (for a similar result cf. also Licht and Steiner (1994)): however, the majority of immigrants in the sample had been in Germany for many years before the start of the panel, so their assimilation could have already been completed; in addition, guest workers who spend more years in Germany are likely to be negatively selected, as successful workers tend to return to their home country while unsuccessful ones remain waiting for an opportunity.

More recently, Chiswick et al. (2005) use three waves of the Longitudinal Survey of Immigrants to Australia to find a negative relation between wage recorded in the first and last wave, which they interpret as sign of higher investment in human capital specific to the country of destination (cf. Duleep and Regets (2002) for theoretical support of such claim); moreover, they find strong inertia in immigrant wages, a sign of the importance of individual fixed effects for economic success.

With the recent surge in availability of administrative panel data, assimilation has been tested with longitudinal data from an increasing number of countries.

Beenstock et al. (2009) use a matched sample from 1983 and 1995 Censuses of Israel, finding that large recent migration inflows (linked to the Post-Soviet *aliyah*) raised the returns to country-specific capital among established immigrants, who experienced a steeper increase in earnings. They also compare longitudinal and cross-sectional analysis results, which seem to contradict each other. Such apparent contradiction is explained by the analysis of sample attrition, as survivor bias is shown to distort cross-sectional estimates due to positive selection of stayers, and the dynamic behaviour of returns to destination-specific skills, which makes the assimilation curve shift over time.

Analysing the impact of remigration on cross-sectional estimates using of social Security records, Lubotsky (2007) finds that the rate of assimilation is much lower than those in the cross-sectional literature, which he (as Beenstock et al.) argues are significantly overestimated due to emigration of unsuccessful immigrants.

For the United Kingdom there exist two panel data studies of assimilation, both employing the Lifetime Labour Market Database (which spans from 1978 to 2006): Dickens and McKnight (2008) and Lemos (2013). Both studies report a fall over the last decades in the immigrant wage penalty at market entry, but they do not agree on convergence rate dynamics: the earlier study reports no significant change in assimilation over the last 40 years (convergence takes around 20 years for male immigrants), while the latter argues that newer immigrants have experienced faster assimilation (while convergence took 30 years for post-war immigrants, the amount is halved for more recent immigrants).

3 Data

3.1 The UK Household Longitudinal Study

I make use of waves 1-5 of the End User Licence[3] (EUL) version of *Understanding Society*: The UK Household Longitudinal Study, a survey that has been conducted by the Institute for Social and Economic Research (ISER) of the University of Essex since 2009 and, due to funding recently received by the Economic and Social Research Council, is currently set to run at least until 2020. *Understanding Society* superseded the 1991-2009 British Household Panel Survey (BHPS) as its enlarged (in sample size terms) and revamped (in terms of questionnaires included) successor.

The peculiarity of *Understanding Society* – and, previously, the BHPS– is that, as opposed to many other datasets available for the United Kingdom, it is not a pseudo-panel of longitudinally independent interviews but a panel that examines an expanding number of households (30169 in wave 1) through time – given that the households remain available for interview. While the sample in itself is unrepresentative, households interviewed are assigned a number of longitudinal and cross-sectional weights for the different questionnaires they respond to. Such weights, which derive the probability of being interviewed based on a multitude of characteristics, can then be included in statistical analysis to correct for sample selection[4].

The survey is composed of questionnaire modules[5], half of which allocated annually and half containing questions asked at larger intervals (for example, some modules have

[3] The End User Licence differs from the richer Special Licence (SL) version of the dataset in that it does not include sensitive data. For example, the SL dataset includes uncapped income, full occupational coding, month of birth; it also includes cross-walk household identifiers to administrative data. However, access to SL data requires a rather lenghty process of application –without adding information crucial to my study that I could not obtain from the EUL dataset.

[4] I choose not to employ weights in the main analysis – please see Appendix B for an explanation. Also cf. Longhi and Rokicka (2012) on representativeness issues of larger traditionally employed datasets, such as Labour Force Survey and Worker Registration Scheme, which do not provide weights.

[5] Available at https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires

only been proposed once, some others are asked every two-three years). Most of the interviews are conducted by computed-assisted personal interviewing (CAPI), with some self-completion modules being answered directly by interviewees on the interviewers' laptops.

Table 1: Longitudinal re-interview rates for adults with full interview in the preceding wave

	Wave 1[a]	Wave 2	Wave 3	Wave 4	Wave 5
Full Interview	81.8 %	72.4 %	78.8 %	82.7 %	85.2 %
N	n.a.	45836	40666	38483	37191
Proxy Interview	5.3 %	1.8 %	2.2 %	2.0 %	1.9 %
N	n.a.	1146	1115	920	841
Telephone Interview[b]		0.5 %			
N		300			
Other non-Interview	6.3 %	1.1 %	1.1 %	0.8 %	0.7 %
N	n.a.	662	571	392	291
Refusal	6.7 %	0.8 %	1.0 %	0.7 %	0.7 %
N	n.a.	527	483	324	314
Household non-Contact		4.9 %	2.7 %	2.4 %	2.1 %
N		3092	1388	1095	
Household Refusal		10.5 %	7.8 %	6.5 %	5.0 %
N		6633	3997	3017	2181
Household Other non-Interview		0.2 %	0.8 %	0.7 %	0.6 %
N		123	409	326	256
Household Untraced		5.9 %	4.1 %	3.1 %	2.8 %
N		3744	2103	1464	1213
Household Ineligible		1.9 %	1.7 %	1.1 %	1.0 %
N		1222	862	533	468
N individual responders	50199	63285	51594	46554	43676

Data from tables 2, 6, 9, 12, 15 in the Understanding Society User Manual.
[a] For Wave 1, only percentages are reported by the Manual.
[b] Telephone Interviews were conducted only for BHPS respondents in Wave 2.

Figure 1: Timing of data collection

Year	Quarter	Survey				
2008	Q1					
	Q2					
	Q3	BHPS Wave 18				
	Q4					
2009	Q1	Wave 1, year 1				
	Q2					
	Q3					
	Q4					
2010	Q1	Wave 1, year 2	Wave 2, year 1*			
	Q2					
	Q3					
	Q4					
2011	Q1		Wave 2, year 2	Wave 3, year 1		
	Q2					
	Q3					
	Q4					
2012	Q1			Wave 3, year 2	Wave 4, year 1	
	Q2					
	Q3					
	Q4					
2013	Q1				Wave 4, year 2	Wave 5, year 1
	Q2					
	Q3					
	Q4					
2014	Q1					Wave 5, year 2
	Q2					
	Q3					
	Q4					

* BHPS becomes sample component in Wave 2, year 1

Interviews in each wave take place over 24 months (figure 1), so that there in no exact wave-year correspondence: however, date of interview is reported for each observation and interviews are organised so that each participating household is interviewed yearly. Individuals who took part in the BHPS are given the choice whether to opt out from the new survey, and if willing to be interviewed under the new survey they are included from the second wave onwards.

As table 1 shows, re-interview rates are high, allowing me to restrict the analysis to adults interviewed for all of the five waves and still obtain a sizeable sample. However, dropping individuals faces me with the problem of non-random sample attrition that might emerge from refusals – both at the household and individual level –, ineligibility and non-traceability. Refusals might be connected with cultural background and socio-economic conditions, possibly leading to under-sampling of poorer households unwilling to disclose sensitive information. Non-natives might be significantly more reluctant to be interviewed and might feel less at ease around British interviewers. Non-traceability, on the other side, is not distinguished by cause: as a result, very different reasons (e.g. a non-communicated change of address; remigration to country of origin) that inhibit household traceability are grouped under a unique header. Similarly, such different events as the communicated emigration of an household and the death of all its members are registered under ineligibility. Notwithstanding the legitimacy of the concerns regarding attrition, one can be reassured of the limited impact of non-traceability, ineligibility and refusal, as the respective rates are significantly low throughout waves. Refusals steadily falling between waves are in particular a sign that in the near future attrition might become negligible.

3.2 From the Download to Statistical Analysis

The survey nature of *Understanding Society* and the breadth of interest of the information it contains impose some dataset manipulation. In particular, such manipulation takes the duplex form of subsampling, i.e. selecting individuals with suitable characteristics among the interviewed population, and derivation, i.e. obtaining relevant information not directly present in the modules – but imputable from related questions – and correcting survey data noise.

3.2.1 Subsampling

As noted, I chose to limit my analysis to individuals observed throughout waves 1-5. Resorting to a balanced panel saves us from having to model panel attrition, which is complex as causes of attrition are for the major part undistinguishable; in addition, a thorough analysis of non-random sample attrition in *Understanding Society* would require a study on its own[6]. In addition, I drop any observation that lacks central information (e.g. labour income, year of arrival for immigrants).

Given that I focus on the analysis of labour market outcomes, as a third step I restrict the sample to individuals between the age of 18 – school leaving age in the UK – and 65 – which during the first waves of the survey was the Default Retirement Age (DRA), the age at which employers could force retirement for workers. Although DRA was abolished in 2011, the choice of 65 as cut-off age is maintained to follow an established convention in labour research. Further, I restrict the sample to individuals who had completed education in 2009: this excludes university and further education students that enter the market in waves other than the first.

These multiple rounds of sample restriction lead me to analyse the labour market dynamics of 14188 individuals (for a total of 70940 individual-wave observations), of which 2344 are foreign-born and 514 are post-2004 immigrants: Table 3 reports summary statistics[7] for the resulting sample. Interestingly, immigrants entering the UK after 2004 are more likely to have obtained a degree than those who migrated before the EU Enlargement, with the latter group reporting on average higher income (accompanied by higher variance) than natives while having comparable age and potential experience. Employment rates are the lowest among Commonwealth immigrants, who perform significantly worse than their non-Commonwealth counterpart and natives.

3.2.2 Variables, Derivation and Noise Correction

Although most of the information I make use of is readily available from the downloadable version of the survey, some variables used in the earnings and employment

[6] cf. Fitzgerald, Gottschalk, Moffitt (1998) for a thorough analysis of attrition in the case of the Michigan Panel Study of Income Dynamics. As for *Understanding Society*, Lynn et al. (2012) is the only piece of research – purely descriptive – that specifically refers to attrition, and is only limited to the first two waves.

[7] Time-dependent values are taken from the first wave.

assimilation specifications have to be derived. For example, potential experience is calculated as years elapsed since leaving education; similarly potential UK experience is the number of years since migration (or, if one has emigrated before completing education, years since leaving education) and potential Home experience the time spent in the country of origin between leaving education and emigrating to the UK. These variables clearly represent a proxy for real labour market experience, as they will overstate true experience in the presence of unemployment spells and ignore experience acquired in parallel with education. If for example some emigrate in response to a lengthy post-education unemployment spell in the country of origin, while others migrate as results of a promotion on the job, $homexp_i$ will not allow to differentiate between the two. One can anyway expect such extreme cases to form a minority of cases, with a much larger portion of immigrants being somewhere in between and thus reducing proxy bias[8].

As normal with surveys, *Understanding Society* presents some noise and individual-level incoherences which, although often hard to tell apart, face the researcher with the necessity of arbitrary decisions. This has been particularly true for the definition of the employment dummy variable, arguably the single most important piece of information one needs in assimilation studies. I report a deeper discussion of the process that led to the final iteration of $employ_{it}$ in Appendix C.2.

The resulting employment rates are reported in Table 2. Sample employment rates are consistently higher than LFS estimates by 3-5%. Following the modifications to *employ*, the number of inconsistencies largely fell. While some noise persists, especially as a result of trying to recover information from variables related to employment and benefits, its entity should not be such as to bias results.

[8]Both Friedberg (2000) and Bell (1997), who divide labour market experience in pre- and post-migration, do not report whether such information is imputed or obtained from the dataset. However, it seems much likelier that the division is the result of a process not unlike mine.

Table 2: Sample and Labour Force Survey employment rates, by sex

	Wave 1 (Jan[a] 2009)	Wave 2 (Jan[a] 2010)	Wave 3 (Jan[a] 2011)	Wave 4 (Jan[a] 2012)	Wave 5 (Jan[a] 2013)	**Total**
Sample						
Men	80.87%	80.79%	80.52%	80.48%	80.74%	80.68%
Women	69.72%	68.60%	68.28%	68.49%	69.25%	68.87%
Total	74.57%	73.90%	73.60%	73.70%	74.25%	74.00%
L F S						
Men	77.29%	74.92%	75.66%	75.60%	75.96%	75.89%
Women	66.16%	65.56%	65.53%	65.45%	66.47%	65.83%
Total	71.67%	70.20%	70.55%	70.48%	71.18%	70.82%

[a] For Labour Force Survey Estimates I make use of January-March Averages

Table 3: Summary Statistics

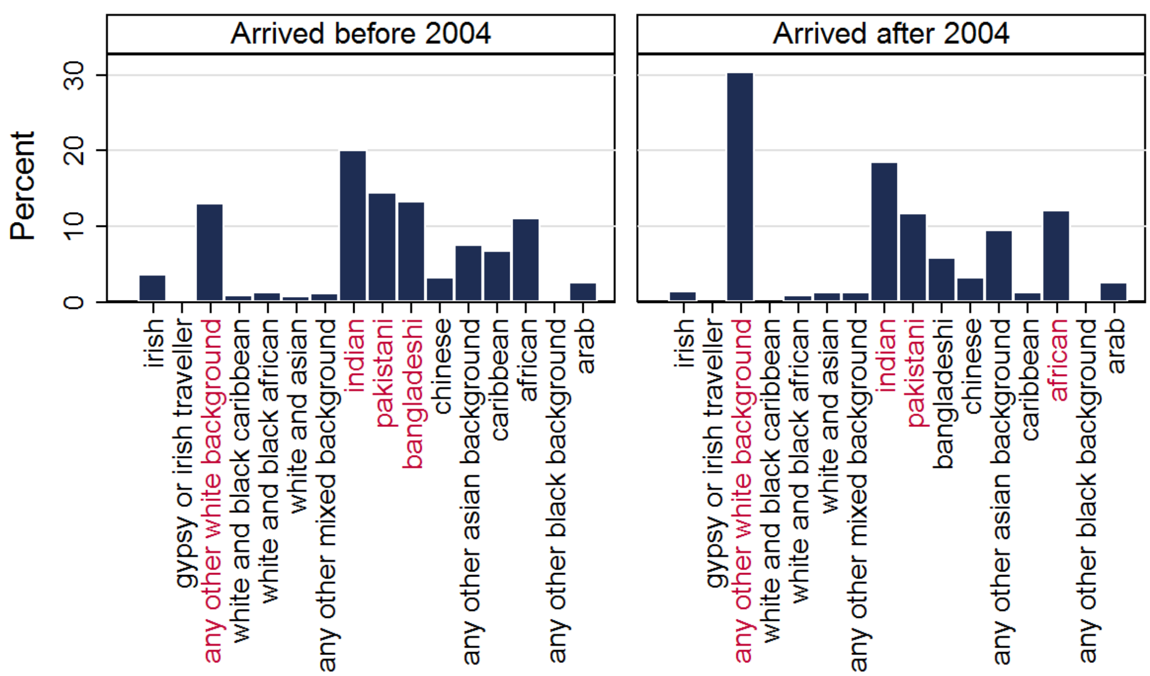
	Natives	Pre-Enlargement immigrants		Post-Enlargement immigrant	
		Commonwealth	Non-Commonwealth	Commonwealth	Non-Commonwealth
Male (*percent*)	44%	44%	40%	48%	42%
Age	44.52	44.44	44.05	35.74	35.66
	(.11)	(.30)	(.37)	(.52)	(.51)
Years Since Arrival	*n.a.*	25.38	24.82	4.98	5.00
		(.40)	(.50)	(.14)	(.13)
Married (*percent*)	55%	73%	59%	86%	63%
Graduates (*percent*)	32%	35%	47%	62%	50%
Employed	76%	61%	74%	65%	70%
Gross Monthly Income:					
Graduates	2772.35	2840.42	2804.96	1973.34	2355.55
	(34.43)	(133.02)	(116.20)	(125.45)	(204.24)
Non-Graduates	1651.81	1504.81	1731.38	1305.87	1414.81
	(16.42)	(52.72)	(74.07)	(105.53)	(90.27)
Potential Experience	24.56	24.43	22.40	14.23	14.34
	(.11)	(.33)	(.41)	(.54)	(.54)
Potential Home Experience	*n.a.*	3.90	3.87	9.32	9.43
		(.18)	(.23)	(.50)	(.52)
N	11,839	798	1,035	246	268

Notes: Statistics taken from the first wave. Standard Errors in parentheses.

3.3 Main Descriptive Statistics

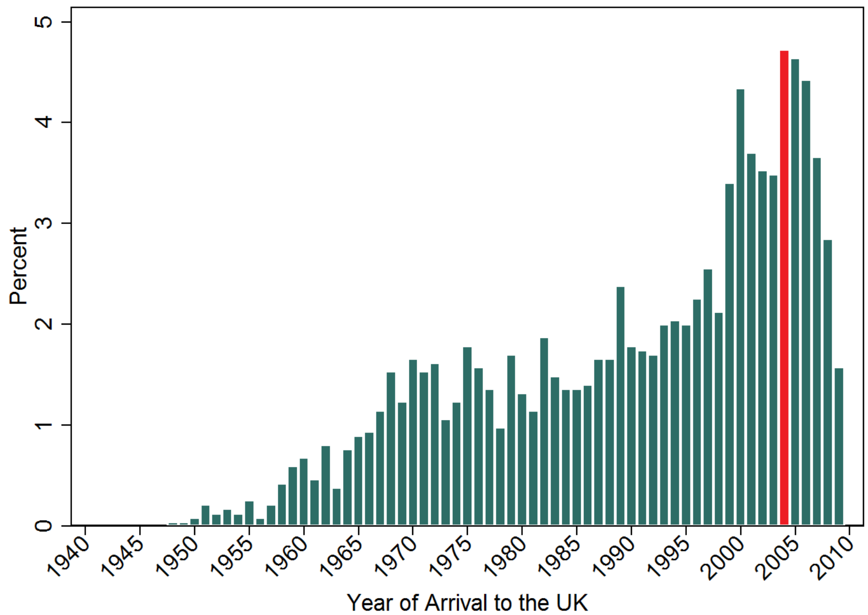
Figure 2 reports the ethnic distribution of immigrants, showing that the shift in origin reported in official figures is realised also in the sample. Other whites – which includes Eastern European Immigrants – rise to being the most important ethnicity after 2004, overtaking Commonwealth countries. Figure 3 shows how the migrants in the sample are distributed along year of arrival. With post-war migration waves aging and being smaller in nominal terms, they became an ever-decreasing portion of migrant labour force. Shares steadily rise until the early 2000s and peak in 2004, the single year with the highest percentage of migrants in our sample. While relative importance of more recent years could be simply due to return migration yet to fully realise, this explanation is not consistent with lower importance of years after 2004[9].

Figure 2: Most represented ethnicities, by Arrival



[9]Comparison with Figure 1 in Dustmann Fabbri (2005) is instructive: it shows a largely similar distribution, even though their sample is the 2004 Labour Force Survey.

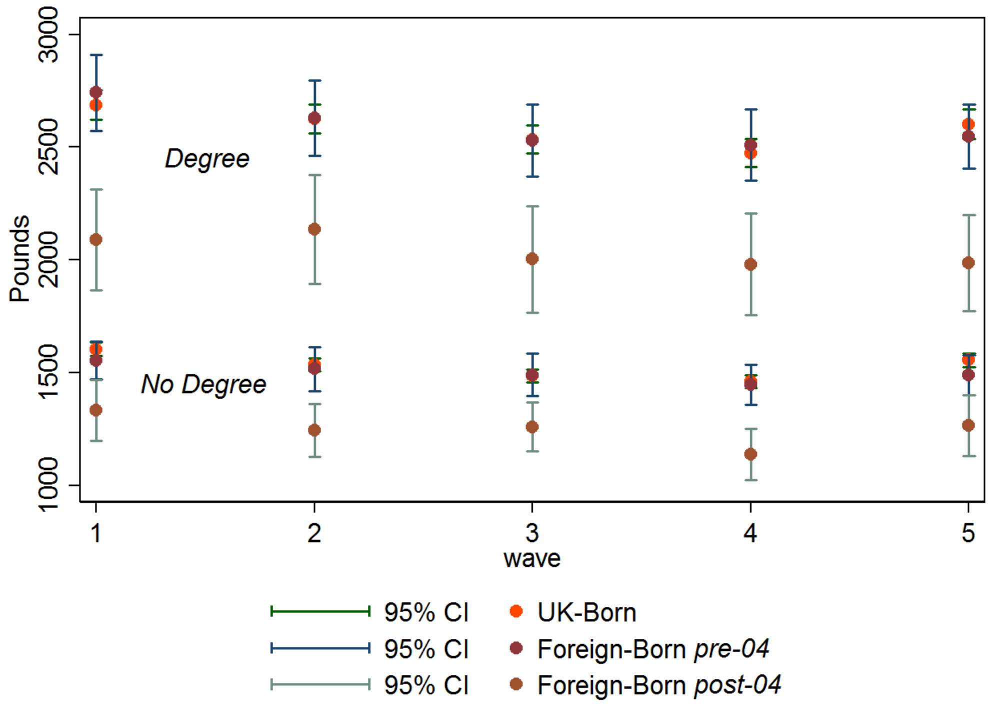
Figure 3: Share of immigrants in the sample, by year of arrival



Note: red bar highlights 2004.

Figure 4 presents a summary of labour net income for six groups: natives, older and newer immigrants with tertiary certification and without. In general, labour income is significantly lower among lower-education workers and it does not seem to grow along waves – if anything, there seems to be a minor U-shaped trend, with income falling in the second and third wave to recover in the later ones. This can be fully reconciled with the Recession and its aftermath: while year-specific effects can be controlled for in estimation to obtain growth rates in absence of the depressive effect of the downturn, mere average income levels will appear stagnant or falling. Comparing trends and levels among groups, older immigrants' dynamics are impressively similar to that of natives; post-Accession immigrants earn significantly less than the other groups, regardless of educational attainment. This gap can be understood as either relative novelty to the British labour market (Chiswick) or lower "quality" of more recent immigrant waves (Borjas). While the two explanations are here observationally undistinguishable and

Figure 4: Income Mean and Standard Deviation, by Degree Ownership and Arrival



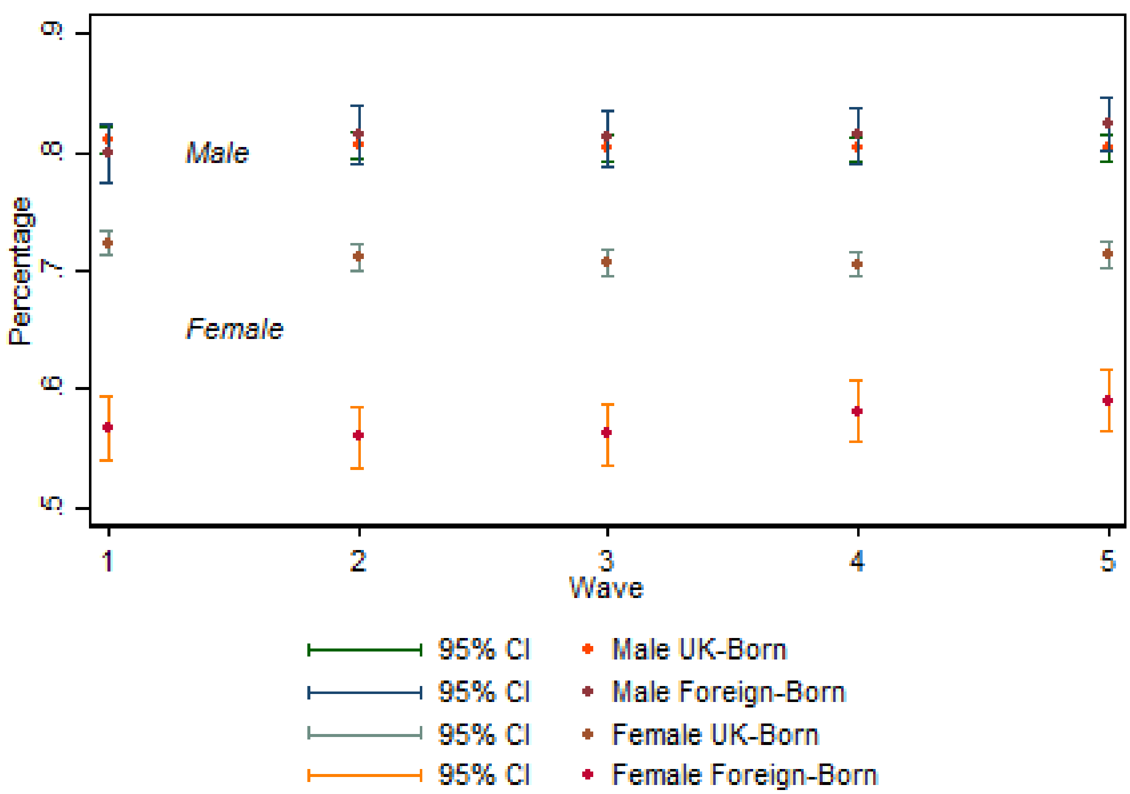
likely to concur, longitudinal analysis will solve the ambiguity through the inclusion of cohort fixed effects.

In Figure 5 I further divide Table 2 employment rates among immigrants and natives. Male employment rates tend to be similar independent of the origin, while native and foreign women appear to have significantly different probability of being employed. This gap is widely acknowledged in research (cf. Clark Drinkwater (2008)) and official releases: elaborating 2011 Census data, ONS (2014)[10] presents employment of 16-64 year old British women close to 70%, with no other non-White ethnic group having similar rates; quarter rolling averages obtained from LFS data (in ONS (2015)[11]) put UK female employment rate at around 75%, with high heterogeneity among immigrant groups. Due to the small samples that would result from a further division, I do not

[10]Available at goo.gl/JP6dOj, point 5 (*link shortened*).
[11]Available at goo.gl/Kfy513, point 6 (*link shortened*).

distinguish between origin of immigrants as done in studies employing larger datasets. Further division[12] by Degree achievement of the four groups unsurprisingly finds that higher level of unemployment among foreign women is connected to their significantly lower number of tertiary-educated individuals; similarly, among immigrant men the lowest levels of employment are to be found among those without tertiary education. The employment gap among education levels does not seem to be as pronounced for natives, the reasons for this asymmetrical behaviour potentially being multiple[12].

Figure 5: Employment rates, by Origin and Sex



[12] Graphs included in Appendix A.

[12] Among others, higher supply of non-graduates allowing employers to favour British natives; higher importance of networks and acquaintance with institutional support –that immigrants might lack – in obtaining low-skill-requirement employment.

4 Methodology

4.1 Earnings Assimilation

Borjas (1999), in the *Handbook of Labor Economics* (ch. 28, pg. 1722), notes: "[...] The choice of standardizing variables is not discussed seriously in most empirical studies in labor economics, where the inclusion criteria seems to be determined by the list of variables available in the survey data under analysis. But this issue plays a significant role in the study of immigrant wage determination [...]". Almost two decades on, the issue still persists: of all the studies I analysed, it is hard to find two that employ the same framework, even though they are mostly adaptations of the original Chiswick (1978) specification. My specification draws mostly from equations (1) and (7) in Friedberg (2000), which I augment to suit my interest.

I first estimate the following equation:

$$\begin{aligned} ln(wage_{it}) = \beta_0 + \lambda_i X_i + \beta_1 exp_{it} + \beta_2 exp_{it}^2 + \beta_3 YSA_{it} + \beta_4 YSA_{it}^2 + \\ + \gamma_1 M_i + \gamma_2 M_i * Post04_i + \alpha_t + \theta_i + \epsilon_{it} \end{aligned} \tag{1}$$

where X_i is a vector of personal characteristics, exp_{it} is potential labour experience (the number of years since the end of studies), YSA_{it} is the number of years since migration, M_i is a dummy that takes value 1 for immigrants, $Post04_i$ a dummy that takes value 1 for immigrants that arrived in or after 2004 and α_t, θ_i are respectively time and individual fixed effects – with $E(\theta_i) = 0$ and $Var(\theta_i) = \sigma_\theta^2$ assumed. Including squared terms of exp_{it} and YSA_{it} accounts for non-linear (expectedly diminishing) returns to the factors.

Given that the dataset provides information regarding year of arrival to the UK and education leaving age, I derive three further variables, $ukexp_{it}$, $homexp_i$ and $expInt_{it}$ – respectively potential UK experience, potential Home experience and the interaction of the two: the coefficient attached to $homexp_i$ allows to estimate returns to pre-migration labour market experience, while that of $expInt_{it}$ should display any complementarity among the two (Bell (1997); Friedberg (2000):

$$
\begin{aligned}
ln(wage_{it}) = \beta_0^* + \lambda_i^* X_i + \beta_1^* ukexp_{it} + \beta_2^* ukexp_{it}^2 + \beta_3^* homexp_i + \beta_4^* homexp_i^2 + \beta_5^* expInt_{it} + \\
+\beta_6^* YSA_{it} + \beta_7^* YSA_{it}^2 + \gamma_1^* M_i + \gamma_2^* M_i * Post04_i + \alpha_t^* + \theta_i^* + \epsilon_{it}^*
\end{aligned}
\tag{2}
$$

Using equation (1) for simplicity and consistency with Borjas (1999), assimilation is defined as

$$
\frac{\partial ln(wage)_{it}}{\partial t}|_{immigrant} - \frac{\partial ln(wage)_{it}}{\partial t}|_{native} = \beta_1^i + 2\beta_2^i exp_{it}^2 + \beta_3 + 2\beta_4 YSA_{it} - (\beta_1^n + 2\beta_2^n exp_{it}^2)
\tag{3}
$$

which simplifies to

$$
\frac{\partial ln(wage)_{it}}{\partial t}|_{immigrant} - \frac{\partial ln(wage)_{it}}{\partial t}|_{native} = \beta_3 + 2\beta_4 YSA_{it}
\tag{4}
$$

as long as $\beta_1^i = \beta_1^n$ and $\beta_2^i = \beta_2^n$. While some foreign studies negate the equality between equations (3) and (4), finding significantly different returns to experience (or age) between the two groups, I employ it as I find no evidence of such difference[13].

[13] Including $forexp_{it}$, the interaction between M_i and exp_{it}, and its square in equation (1) yields consistently insignificant differences – cf. Appendix D.1 for regression output and discussion.

I estimate the parameters in equations (1) and (2) in three main ways: first – ignoring the longitudinal nature of the data – by Pooled Ordinary Least Squares (from here onwards POLS); secondly, modelling individual effects as Random Effects, I make use of the Within Group (WG) and Feasible Generalised Least Squares with Mundlak (1978) correction (FGLS) estimators to exploit the longitudinal dataset[14].

In the discussion I will focus on FGLS estimates, reporting WG and POLS estimates in Appendix D.2. The POLS estimator is in fact inefficient whenever $Var(\theta_i) = \sigma_\theta^2 \neq 0$, being distributed as

$$\hat{\beta}_{POLS} \sim AsyN(\beta, \big(\sum_i X_i'X_i\big)^{-1}\big(\sum_i X_i'\Omega X_i\big)\big(\sum_i X_i'X_i\big)^{-1}) \tag{5}$$

with $\Omega = (\sigma_\epsilon^2 + \sigma_\theta^2)I_t$. Nevertheless, its estimates can be interesting to roughly assess the magnitude of bias associated with treating longitudinal data as cross-sectional, as done in pseudo-panel studies.

The WG estimator solves the inefficiency of POLS by transforming the data in a way that forces the loss of significant information. It in fact involves calculating

$$y_{it} - \bar{y_i} = (x_{it} - \bar{x_{it}})'\beta + (\epsilon_{it} - \bar{\epsilon_i}) \tag{6}$$

which nets out the effect of time-invariant ($\bar{\theta_{it}} = \bar{\theta_i} = \theta_i$) individual heterogeneity, allowing the transformed equation to be estimated by OLS. Along with heterogeneity, however, WG eliminates any other variable that lacks within-individual variation: estimating coefficients of M_i, $Post04_i$, region of origin and others is not possible by WG.

[14] I avoid making use of another estimator popular in the labour literature, the Arellano Bond (1991) estimator, as the inclusion of lagged values of $ln(wage)_{it}$ makes it hard to disentangle their effects from returns to exp_{it} (cf. Cameron Trivedi (2010) pg. 298) and I prefer focusing on the latter as they are central to the measurement of assimilation.

In principle, the standard version of the FGLS estimator allows to both account for the composite error structure and to estimate coefficients of time-invariant regressors, as it implies estimating the variance matrix in a first step. This comes at the cost of assuming θ_i to be uncorrelated with regressors – too strong an assumption in this case, as confirmed by the Hausman[15] test. However, one can model θ_i by specifying its relationship with x_{it} : in particular, also given the short time-span of the panel, we can assume (as suggested in Mundlak (1978))

$$\theta_i = \bar{x}_i{}'\delta + v_i \qquad with \qquad v_i \perp \bar{x}_i \, by\, construction;\; E(v_i) = 0 \tag{7}$$

which will pick up any linear relationship between covariates and error term, allowing for consistent estimation. We thus obtain the FGLS estimator with Mundlak correction[16]

$$y_{it} = x'_{it}\beta + \bar{x}_i{}'\delta + v_i + \epsilon_{it} \tag{8}$$

To answer possible concerns over within-autocorrelation (i.e. at the individual level) of the error term ϵ_{it} (or ϵ^*_{it}), I make use of the serial correlation test developed in Wooldridge (2010) (pg. 320, see also Drukker(2003)). Such test stores the residuals from a regression of first-differences to test whether $\text{Corr}(\Delta\epsilon_{it}, \Delta\epsilon_{it-1}) = -0.5$, as should be if the error term is not serially correlated: hypotheses of autocorrelation are consistently rejected across specifications of the test, reported in Appendix D.6.

In Appendix D.6 I also include estimates from a Baltagi Wu (1999)-transformed Mundlak FGLS estimator, which accounts for the AR(1) component in the data allowing then for efficient estimation. Coefficients are, as expected, broadly consistent with Table 4.

[15] cf. Appendix D.5.

[16] This is implemented in STATA by including, in addition to x_{it}s specified in equations (1) and (2), within-individual average of time-variant regressors. T-testing coefficients attached to individual means can also serve as an additional test of fixed vs. random effects: cf. Appendix D.5.

4.2 Employment Probability Assimilation

Symmetrically to the analysis of earnings assimilation, I estimate the following equation for employment:

$$employ_{it} = \Lambda\big(\xi_0 + \chi_i X_i + \xi_1 exp_{it} + \xi_2 exp_{it}^2 + \xi_3 YSA_{it} + \xi_4 YSA_{it}^2 + \\ +\mu_1 M_i + \mu_2 M_i * Post04_i + \alpha_t + \theta_i + \nu_{it}\big) \tag{9}$$

and

$$employ_{it} = \Lambda\big(\xi_0^* + \chi_i^* X_i + \xi_1^* ukexp_{it} + \xi_2^* ukexp_{it}^2 + \xi_3^* homexp_i + \xi_4^* homexp_i^2 + \\ +\xi_5^* YSA_{it} + \xi_6^* YSA_{it}^2 + \mu_1^* M_i + \mu_2^* M_i * Post04_i + \alpha_t^* + \theta_i^* + \nu_{it}^*\big) \tag{10}$$

in which evidence of assimilation is extrapolated from, as in section 4.1, ξ_3, ξ_4, ξ_5^* and ξ_6^*. In equations (9) and (10) the binary variable $employ_{it}$ is thought of as the result of a latent variable threshold model, that is

$$employ_{it} = \mathbf{1}[y_{it}^* \geq 0] = \mathbf{1}[x_{it}'\xi + \theta_i + \psi_{it} \geq 0] \quad as \;\; y_{it}^* = x_{it}'\xi + \theta_i + \psi_{it} \tag{11}$$

where $\mathbf{1}$ is an indicator function taking value 1 when the event in parentheses is realised (0 otherwise) and ψ_{it} is $iid \sim \Lambda(0, \frac{\pi^2}{3})$. The Random Effects Logit model specifies

$$Pr(employ_{it} = 1 | x_{it}, \xi, \theta_i) = \Lambda(x_{it}'\xi + \theta_i) \qquad with \qquad \Lambda(x_{it}'\xi) = \frac{e^{x'\xi}}{1 + e^{x'\xi}} \tag{12}$$

and $\theta_i \sim N(0, \sigma_\theta^2)$ as before. The joint density for individual i then is, integrating θ_i out (Cameron Trivedi (2010), pg. 625; Wooldridge (2010), pg. 613)

(13)

$$f(employ_{it},\dots,employ_{iT}) = \int_{-\infty}^{+\infty}\left[\prod_{t=1}^{T}\Lambda(x_t'\xi+\theta_i)^{employ_t}\{1-\Lambda(x_t'\xi+\theta_i)^{1-employ_t}\}\right]g(\theta_i|\sigma^2)d\theta_i$$

where $g(\theta_i|\sigma^2)$ is the density of $N(0,\sigma_\theta^2)$. Having no analytical solution, integral (13) is calculated through numerical methods in STATA – through adaptive 12-point Gauss-Hermite quadrature (as first suggested by Butler and Moffitt (1982)).

Mundlak correction is included for the reasons discussed in subsection 4.1 (cf. Greene (2012), pg. 767; Wooldridge (2010), pg. 615). It is an improvement over "naïve" Random Effects Logit as it allows for a more general $\theta_i|x_i \sim N(\bar{x_i}'\xi,\sigma_\theta^2)$, yielding

(14) $$Pr(employ_{it}=1|x_{it},\xi,\theta_i) = \Lambda(x_{it}'\xi+\bar{x_i}'\delta+\upsilon_i) \qquad as \quad \theta_i = \bar{x_i}'\delta+\upsilon_i$$

however (as noted in Hsiao (2014), ch.7 pg. 244), while decomposing individual heterogeneity as by equation (14) is rather innocuous in the linear case, applying the Mundlak – or, alternatively, the Chamberlain (1980, 1984)[17] – correction to the non-linear case comes at the cost of assuming the relationship between θ_i and x_{it} to be exactly linear, and υ_i to be independent of $\bar{x}_{it}$ with its own defined distribution: I make use of this stricter assumption.

In addition to estimating (9) by Mundlak RE Logit Maximum Likelihood Estimator (MLE), I also use the standard Logit MLE (which ignores θ_i and the longitudinal dimension and, as with POLS, will present biased estimates) and the Conditional Fixed Effects Logit MLE (comparable to WG in that both procedures do not allow to esti-

[17] The Chamberlain (1980, 1984) correction defines the incidental parameter θ_i as a function of all individual x_{it} – i.e. x_i – instead of the average $\bar{x_{it}}$. Its interpretation is fundamentally the same as that of the Mundlak (1978) correction in the linear model; in non-linear models, the two approaches diverge (cf. Cameron Trivedi (2005), pg. 786): I choose to employ Mundlak specification for consistency – it also does not hurt that it implies estimating less parameters in the burdensome *xtlogit* quadrature.

mate the impact of time-invariant regressors), as estimation of non-linear fixed effects models is complex[18].

As integrating θ_i out implies $Pr(employ_{it} = 1|x_{it}, \xi) \neq \Lambda(x'_{it}\xi)$ (Cameron Trivedi (2010), pg. 626), parameters obtained from Logit and Panel Logit models are not directly comparable as are, for example, parameters from POLS and FGLS[19]. Keeping this in mind, from each of the three estimators I estimate Average Marginal Effects (AMEs, also known as Average Partial Effects), i.e. the sample average of marginal effects for every individual observation, as Logit coefficients are not meaningful for analysis purposes.

I believe AMEs to be more informative than what are often considered an alternative, Marginal Effects at Mean (MEMs), as the former averages the marginal effects with regard to x_{it} of individuals in the sample while the latter calculates marginal effects of x_{it} for an individual endowed with sample average characteristics. As also noted in Wooldridge ((2010), ch.2 pg. 22), in cases where the marginal effect

$$\partial E(employ_{it}|x_{it}, \theta_i)/\partial x_{it}$$

is so intrinsically connected with θ_i it makes little sense to average out sample individual characteristics and expect the marginal effect calculated from an imputed 'average' individual to be representative: instead, averaging individual marginal effects, which will take into account incidental heterogeneity, can be of higher interest.

[18] cf. Greene (2004). As Wooldridge (2010) notes for the Probit case, the common misuse of the "*Fixed* Effects" terminology is "unfortunate" (ch. 15 pg. 612): it might in fact induce to believe θ_i is some parameter to be estimated, which can potentially bias results. Fixed Effects rather refers to the idea that we can "consistently estimate the parameters β [ξ in our case] without specifying a distribution for c_i [θ_i] given x_i" (*ibidem*): it is still assumed that θ_i is a pick from a distribution.

[19] One could estimate Marginal Effects at $\theta_i = \bar{x_i}'\delta$, but such value could be 'nonrepresentative' (Cameron Trivedi (2010) pg. 627); in addition, as Wooldridge (2010) notes, when θ_i is continuous there is technically no one endowed with a specific value of θ_i.

5 Results

5.1 Earnings Assimilation

Table 4: Feasible GLS with Mundlak (1978) Correction Estimates of eq's (1) and (2)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-0.387***		-0.286*	-0.112	
	(0.10)		(0.17)	(0.16)	
Post-04 Immigrant			-0.090	-0.091	
			(0.06)	(0.06)	
Potential Experience	0.040***	0.040***	0.040***		
	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.039***	-0.039***	-0.039***		
	(0.00)	(0.00)	(0.01)		
Potential UK Experience				0.040***	0.040***
				(0.00)	(0.00)
Potential UK Experience Squared x0.01				-0.037***	-0.037***
				(0.01)	(0.01)
Potential Home Experience (Immigrants only)				0.008	0.011
				(0.01)	(0.01)
Potential Home Experience Squared x0.01 (Immigrants only)				-0.028	-0.039
				(0.03)	(0.03)
Home and UK Exp Interaction (Immigrants only)				-0.060	-0.058
				(0.07)	(0.07)
Years Since Arrival to the UK	0.027***	0.027***	0.027***	0.023*	0.023*
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK Squared x0.01	-0.055***	-0.055***	-0.055***	-0.039	-0.039*
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
European		-0.221**			-0.065
		(0.09)			(0.13)
Asian		-0.518***			-0.352***
		(0.11)			(0.13)
African		-0.366***			-0.201
		(0.10)			(0.14)
North American		-0.242**			-0.071
		(0.10)			(0.16)
Centre-South American		-0.402***			-0.244
		(0.10)			(0.16)
Australian or New Zealander		0.017			0.194
		(0.10)			(0.16)
Non-specified Country		-0.378***			-0.217
		(0.09)			(0.14)
Observations	47,333	47,333	47,333	47,333	47,333
Number of pidp	11,190	11,190	11,190	11,190	11,190

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Omitted output: Male dummy, Education, Marriage Dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE, Mundlak correction.

Table 5: Allowing for Regional and Post-04 heterogeneity in YSA and YSA^2 estimates.

	Experience Unified			UK/Home Experience	
	(1)	(2)	(3)	(4)	(5)
VARIABLES	YSA by region	YSA by region	YSA by Post-04	YSA by region	YSA by Post-04
Years Since Arrival to the UK			0.034***		0.031**
			(0.01)		(0.01)
Years Since Arrival to the UK Squared x0.01			-0.067***		-0.050**
			(0.02)		(0.02)
YSA – Post-04			-0.078		-0.075
			(0.07)		(0.07)
YSA^2 – Post-04			0.624		0.609
			(0.57)		(0.56)
Immigrant	-0.421***		-0.153	-0.313***	0.035
	(0.12)		(0.16)	(0.11)	(0.17)
Post-04 Immigrant			-0.438		-0.504
			(0.63)		(0.62)
YSA – European	0.042***	0.030***		0.029***	
	(0.01)	(0.01)		(0.01)	
YSA – Asian	0.009	0.012		-0.001	
	(0.01)	(0.01)		(0.01)	
YSA – African	0.031***	0.035***		0.022*	
	(0.01)	(0.01)		(0.01)	
YSA – North American	0.048***	0.036		0.038*	
	(0.02)	(0.03)		(0.02)	
YSA – Centre-South American	0.021	0.030		0.014	
	(0.02)	(0.03)		(0.02)	
YSA – Australian or New Zealander	0.063***	-0.006		0.058***	
	(0.02)	(0.03)		(0.02)	
YSA – Non-Specified Country	0.028	0.032		0.022	
	(0.02)	(0.02)		(0.03)	
YSA^2 – European	-0.082***	-0.063***		-0.051***	
	(0.02)	(0.02)		(0.02)	
YSA^2 – Asian	-0.019	-0.024		0.006	
	(0.02)	(0.02)		(0.02)	
YSA^2 – African	-0.069***	-0.076***		-0.043**	
	(0.02)	(0.02)		(0.02)	
YSA^2 – North American	-0.105***	-0.087*		-0.077*	
	(0.04)	(0.05)		(0.04)	
YSA^2 – Centre-South American	-0.046	-0.059		-0.026	
	(0.03)	(0.04)		(0.04)	
YSA^2 – Australian or New Zealander	-0.133***	-0.027		-0.118***	
	(0.04)	(0.05)		(0.04)	
YSA^2 – Non-Specified Country	-0.064	-0.069		-0.043	
	(0.04)	(0.04)		(0.05)	
European		-0.290**			
		(0.12)			
Asian		-0.484***			
		(0.12)			
African		-0.519***			
		(0.16)			
North American		-0.322			
		(0.28)			
Centre-South American		-0.645*			
		(0.34)			
Australian or New Zealander		0.433			
		(0.44)			
Non-Specified Country		-0.579**			
		(0.23)			
Observations	47,333	47,333	47,333	47,333	47,333
Number of pidp	11,190	11,190	11,190	11,190	11,190

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . Years Since Arrival Squared (YSA^2) are x0.01 where not specified. Omitted output: Male dummy, Experience, Education, Marriage Dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE, Mundlak Correction.

Tables 4, A3 and A4 each present the estimates of the three main estimators I use (FGLS, POLS and WG) for three specifications of the earnings assimilation equation: 'Gap at entry' (columns (1) and (2)) does not differentiate between pre- and post-04 entry, thus being comparable to the specifications employed by many previous studies. 'Post-04 Gap' and 'Dividing Experience' instead refer to equations (1) and (2). In columns (1), (3) and (4) I employ the "immigrant" dummies to estimate average gap at entry; in the other columns I substitute the two with a full set of dummies for the region of provenance of the immigrant (omitting UK – i.e. natives – as term of comparison) to estimate region-specific gap at entry.

In the most naive specification of Table 4, the immigrant deficit at entry is estimated to be around 32.1[20]%: such estimate is not far from that of Lemos (2013)[21] and Dickens McKnight(2008). The gap reduces to 24.9% when immigrants are divided by arrival in column (3) – with newer immigrants not found to be entering the market at an additional disadvantage[22] – and becomes insignificant when I differentiate Home and British potential experience[23]. Returns to pre-migration experience are less than half of those to UK experience and insignificant, a sign of the very limited transferability of both its skill and signal content consistent with evidence of over-skilling among British immigrants[24]: over-skilling might in fact be educational but also professional.

[20]It has become standard practice in the economic literature to ignore that $\%\Delta y = 100[e^{\hat{\beta}} - 1]$ and make use of the sheer log-level regression coefficients when reporting percentage changes. As the quality of such approximation falls drastically as the coefficient increases, I report the percentage change obtained by using the mentioned formula – which explains the discrepancies with table output.

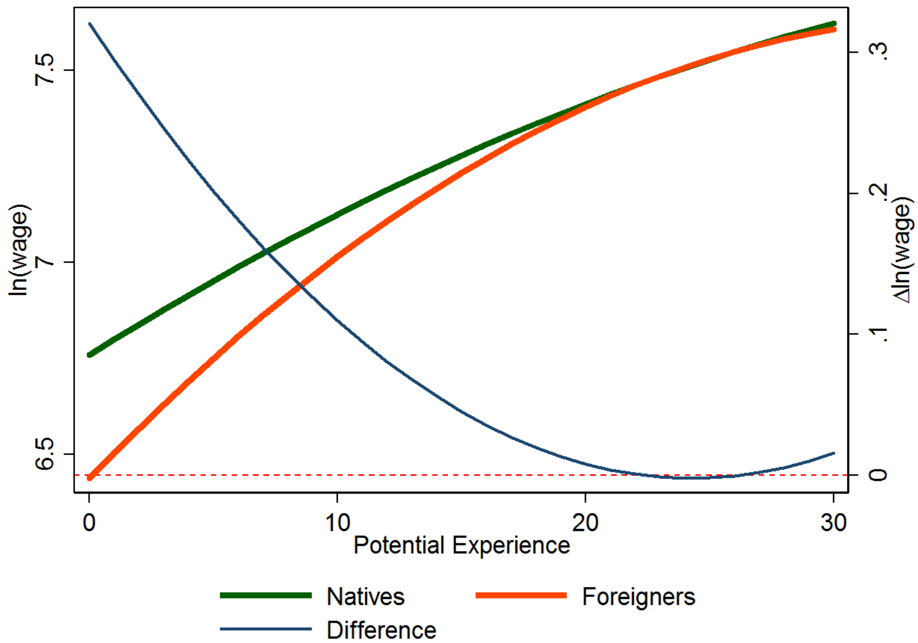
[21]Lemos obtains a coefficient of -0.462 and claims that the gap at entry is 46.2%: using the formula from footnote 20, the gap is in reality of 37% – almost 10% off from the approximation.

[22]Significance of the $M_i * Post04_i$ coefficient in column (4), Table A3 is explained by the group of comparison being the entire population of natives and pre-2004 immigrants.

[23]This reversal is testified even more strongly in Table 4, column (2) of Friedberg (2000) – where dividing experience and education in pre- and post-arrival determines the Immigrant dummy coefficient to turn positive: such result is not discussed in the paper.

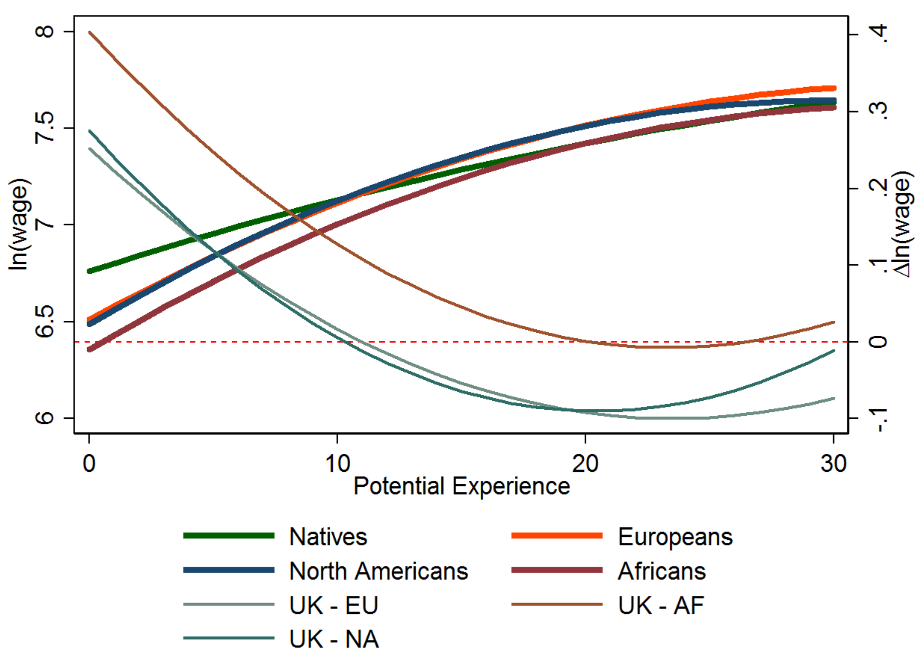
[24]According to the MAC *Migrants in low-skilled work* (2014) report, "around one-third non-EU and EU8 and EU2 workers in low-skilled jobs left full-time education at age 21 or greater, while this is true for only 10 per cent of UK workers in low-skilled jobs" (pg. 119).

Figure 6: Experience–Earnings Profile



Note: Profiles resulting from Table 4, Column 1 coefficients. Difference on the right axis, red dashed line highlights 0 difference.

Figure 7: Experience–Earnings Profiles by Region



Note: Profiles resulting from Table 5, Column 2 coefficients. Profiles drawn only for origin groups that display convergence, i.e. initial disadvantage in Table 4 and significant returns to YSA_{it} in Table 5. Difference on the right axis, red dashed line highlights 0 difference.

Division by region of arrival confirms the heterogeneity highlighted by most of the international literature: Europeans and North Americans display the lowest deficits, while Asians appear particularly disadvantaged and immigrants from Oceania endure no deficit at all.

Returns to experience are in the range of 4%[25] per year, and the (declining) rate of yearly convergence for immigrants is in excess of 2% across most specifications[26]. While there is no agreement within the literature around returns to YSA_{it} – rightfully so, as they are entwined with the labour market of the country, its immigration flows and the sample analysed – my estimate is consistent with Lemos'[27] and more optimistic than Bell (1997), who finds insignificant returns.

The implied period of erosion of the 32.1%[28] deficit at entry can be calculated by solving (4):

$$\frac{\partial ln(wage)_{it}}{\partial t}|_{immigrant} - \frac{\partial ln(wage)_{it}}{\partial t}|_{native} = \beta_3 + 2\beta_4 YSA_{it} = 0.023 - 2 * (0.00044)YSA_{it} \quad (15)$$

yielding $YSA_{it} = 26.14$. Such estimate is fully in line with recent longitudinal research, opposing the finding of Bell (1997) that convergence is not realised within the working life of the immigrant. The observed heterogeneity of initial deficit holds also in the number of years needed for convergence, as shown in Table 5: as claimed already in Borjas (1985), not all immigrant groups assimilate. I find insignificant assimilation for Asian and Centre-American workers – who also experienced the largest entry earnings gap –, while Europeans and North Americans both reach native wage levels in little

[25] cf. Funkhouser Trejo (1995), who for the US find returns in order of 5%, and Pischke (1992) that finds returns in the order of 4% for the German case. Studies that proxy experience by age, such as Lemos (2013), Dustmann Frattini (2005) and, for the US, Borjas(1994) consistently find higher returns to such proxy.

[26] exp_{it} and YSA_{it} coefficients varying significantly among POLS and WG-FGLS serves as a reminder of the significant biases arising from ignoring the longitudinal structure of the sample.

[27] Table A3, column 3. This is the only longitudinal estimate I could find for the United Kingdom: Dickens and McKnight (2008) unfortunately do not report estimated convergence rates.

[28] Table 4, column (1) estimates – the resulting number of years is fairly insensible to the choice.

more than a decade. I plot these assimilation paths graphically in figures 6 and 7. No evidence of assimilation rates differing among pre-2004 and post-2004 immigrant populations[29] is found: this result, along the insignificant additional earnings gap at entry that post-04 immigrants experience, seems to dispel the claim that more recent immigrants perform worse than their established counterpart.

5.2 Employment Probability Assimilation

Table 6 presents a picture of employment that is significantly different from that of earnings. When simply dividing natives and immigrants, neither being foreign-born nor being a recent immigrant are found to be associated with significant lower probability of employment: the only relevant factors appear to be experience and educational attainment (and, in omitted output, being male). Each additional year of experience in the country of origin increases the probability of employment by 1.3%, remarkably more than an additional year of British experience (0.3%): immigrants with previous experience are thus advantaged compared to migrants and natives who enter the labour market for the first time.

Dividing immigrants by region of origin displays a pattern of heterogeneity different from the one found in earnings: Europeans and immigrants from Oceania are more likely to be employed at entry than British with similar specifics by a consistent margin, while Asians show a 8% deficit. Not only, therefore, is the status of immigrant – neither recent immigrant, from a flow that has been argued to be qualitatively inferior to labour arrived to the UK before 2004, nor established immigrant – not consistently connected to a higher level of unemployment, but some groups even display higher chance of being employed – be it for self-selection or because they arrive as result of a job offer.

[29] I omit a comparison of pre- and post-04 immigrants divided by region as Post-04 immigrants are particularly few for Oceania and the Americas. Intersecting both analyses is possible only with larger datasets and should be an objective of future research.

Table 6: AMEs for the Mundlak (1978)-corrected Panel Logit Estimator of (10)

	General			By Origin	
VARIABLES	dy/dx	$P > \|z\|$		dy/dx	$P > \|z\|$
Potential UK Experience	.003	.013		.003	.000
	(.001)			(.001)	
Potential UK Experience Squared x0.01	-.012	.000		-.012	.014
	(.002)			(.002)	
Years Since Arrival	.002	.636		.001	.634
	(.003)			(.003)	
Years Since Arrival Squared x0.01	.001	.909		.000	.926
	(.005)			(.005)	
Potential Home Experience (Immigrants only)	.013	.000		.014	.000
	(.003)			(.002)	
Potential Home Experience Squared x0.01 (Immigrants only)	-.021	.003		-.027	.000
	(.007)			(.007)	
Home and UK Exp Interaction (Immigrants only)	-.015	.449		-.014	.453
	(.020)			(.019)	
Immigrant	-.013	.599			
	(.025)				
Post-04 Immigrant	-.001	.944			
	(.016)				
European				.057	.003
				(.020)	
Asian				-.082	.000
				(.020)	
African				.010	.676
				(.024)	
North American				.026	.356
				(.028)	
Centre-South American				.004	.925
				(.037)	
Australian or New Zealander				.080	.041
				(.039)	
Non-Specified Country				-.013	.552
				(.022)	
No Qualification	*term of comparison*				
Other Qualification	.212	.000		.151	.000
	(.017)			(.013)	
GCSE	.288	.000		.212	.000
	(.015)			(.011)	
A Levels	.356	.000		.273	.000
	(.014)			(.011)	
Degree	.401	.000		.329	.000
	(.014)			(.011)	
N	70,940	70,940	70,940	70,940	70,940

Notes: dy/dx for factor levels is the discrete change from the base level.
Omitted output: Male Dummy, Wave fixed effects, Mundlak Correction.
Region FE, Region of birth FE and Year*Region FE not included as they increase computational time exponentially.

6 Summary and Conclusions

This paper provides an assessment of economic assimilation of migrants to the United Kingdom in terms of earnings level and employment probability. Adding to the historically meagre British literature on the matter, I aim to expand available knowledge on the fortunes of those who arrived after the 2004 EU Enlargement, not analysed before in comparable studies, and update assimilation estimates employing a dataset previously unused for the purpose.

Analysing publicly available data from the first five waves of the UK Household Longitudinal Study, the study finds that the average immigrant entering the UK labour market experiences an initial deficit in the order of 30% *vis-à-vis* comparable natives, and that – with an estimated assimilation rate of over 2% – it takes them more than two decades to bridge the gap. Both the deficit and the rate of convergence present large heterogeneity by region of provenance, with Asians and Latin-Americans performing the worst and immigrants from Oceania presenting no gap at entry. No evidence is found of post-2004 immigrants assimilating more slowly or starting at significantly lower levels of earnings than their established counterpart, against popular claims of falling *quality* among recent comers. Dividing experience on the basis of where it was obtained shows that foreign experience is not significantly rewarded on the market, in accordance with the over-qualification literature finding immigrant education being rewarded significantly less than native.

As for the probability of employment, being foreign-born is not associated with significant differences at entry – and additional years in the UK do not benefit immigrants more than natives: for both groups, the employment probability seems to depend solely on gender, education and experience. The asymmetry of earnings and employment deficit at entry could be reconciled by immigrants accepting jobs they are overqualified for, a possible result of skills mismatch or excess labour supply. This is endorsed by the finding that pre-migration experience is associated with increased probability of employment, but its monetary returns are much lower than returns to British experience. The portability of human capital is thus significant, but partial at best.

These results are particularly interesting in light of documented radical compositional (and qualitative) differences between pre- and post-2004 Enlargement flows: the importance of such changes is mostly replicated in the dataset, and yet the assimilation patterns of the two immigrant populations are not found to be different. Immigration is too often presented in terms of summary statistics, be it stocks or flows: as I show, a structural analysis of the data can present a different picture from a descriptive approach. I acknowledge that the limitations of the sample employed in this study might be such as to make the results disputable, but I believe the main takeaways to be the methodological framing to test the assimilation hypothesis and the possibility for apparently conflicting analyses to coexist, rather than the results from the application. The framework seems in fact sound enough to be employed in the analysis of other British longitudinal datasets, the unavailability of which has hampered the transition away from survey data towards administrative data experienced by major microeconomic studies[30], as well as advances in studies of migrant performance. Increased availability of administrative data would pave the way for a multiplicity of studies and an informed debate on migration.

More research on the assimilation of recent British migrants is direly needed: longitudinal studies have been able to present a superficial picture of immigrant income dynamics, but they are already dated and have limited say regarding recent flows. The most recent study analysed (publicly unavailable) data up to 2006; its closest predecessor did not go past 1992. Any paper that updates our knowledge regarding

[30] cf. Chetty (2012) slides.

the economic convergence path of post-Enlargement migrants represents an important addition to a so-far-unexplored area of study. This dissertation tries to move a step forward towards the creation of a structure for such research.

Appendices

A Further Descriptives

Figure A1: Male Employment, by Birth and Degree Ownership

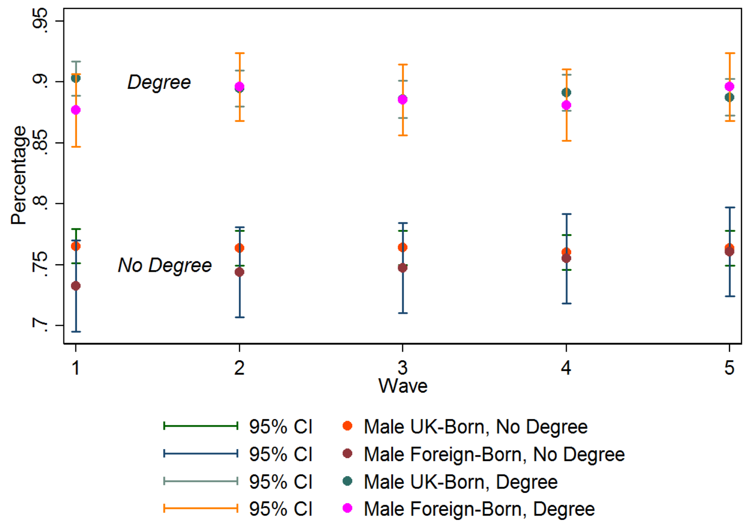
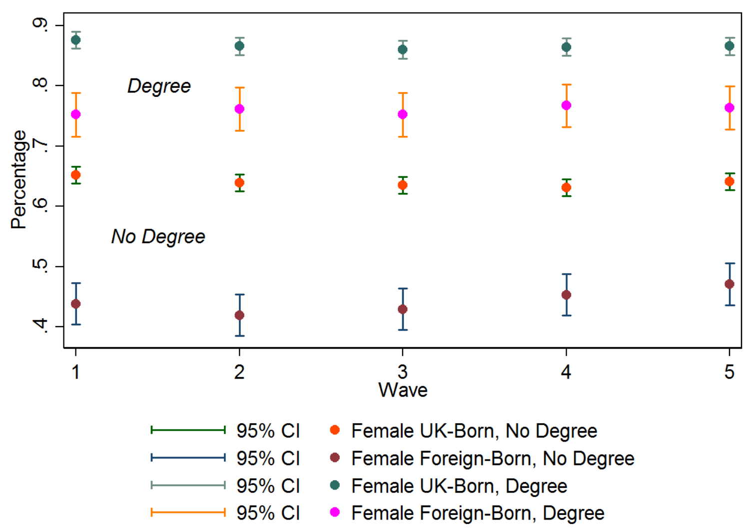


Figure A2: Female Employment, by Birth and Degree Ownership



B Survey Weights: a Note

The ISER provides both cross-sectional and longitudinal sampling weights for the UK Household Longitudinal Survey, along with more information on the structure of data collection.

The use of survey weights in my estimation is problematic for two main reasons: in general, considering only a non-random sub-sample of the surveyed population makes it so that the weights are unlikely to fulfil their intended purpose – and thus their significance is arguable; in addition, STATA does not allow the inclusion of weights for the FGLS estimator.

Regarding the first point, the consensus seems to be that weights ought to be included even when the sample is not considered completely, hoping that selection does not excessively bias the distribution emerging from remaining weights. However, as weights cannot be included in the Mundlak-corrected FGLS estimator, I refrain from using them in Section 5 and related appendices to preserve consistency of output from comparable regressions.

Cameron and Trivedi ((2005), pg. 820) provide further insight on the matter: they claim that weights should be used when using a "descriptive" approach, while a "structural" approach that assumes correct model specification can avoid them. They also refer to the nature and purpose of the survey analysed as a relevant factor for distinguishing the two approaches: large cross-sectional surveys, like the CPS in the US or the LFS in the UK, are conceived for descriptive analysis at national and regional level; other minor surveys – like *Understanding Society* – "are developed with a structural approach in mind" (*ibidem*): that is, they rely on correct model specification by the researcher, rather than on full representativeness, to yield sensible estimates. The authors make the example of determining returns to schooling: if one tries to do so regressing earnings on education and socio-economic characteristics by

OLS, the inclusion of weights is sensible as it avoids unrepresentativeness of the sample in establishing the connection – merely descriptive – between earnings and schooling. If one instead makes use of more advanced estimation methods (instrumental variables, panel data estimators) that allow for causal inference, the introduction of weights is beneficial as it improves efficiency, but not crucial. This is because weighting has the purpose of replicating the real distribution of characteristics and related heterogeneity that in standard least-squares estimation can influence the relationship analysed, but that other techniques can account for.

Ultimately, the sample employed in the study is clearly not representative of the British working population. However, while including weights could be an improvement, analysing a subsection of the population for which the weights have been designed could bias the distribution resulting from remaining weights. In addition, a compelling argument can be made for the minor bias resulting due to the avoidance of weights when the purpose of the study is analytical rather than descriptive and more advanced estimation procedures are employed. For these reasons, although I report estimates of POLS and WG including weights in Appendix D.4, I place unweighed estimates at the centre of my analysis.

C Information on Variables Used and Constructed

C.1 List of Variables Employed in Regressions

varname	**Name**	**Details**
pidp	Personal ID	The identifier that allows to connect observations longitudinally.
wave	Survey Wave	A wave identifier created when merging different waves into a unique dataset.
yearinterview	Interview Year	Year component of the date of interview.
male	Male Dummy	=1 if the respondent is male.
married	Marriage Dummy	=1 if the respondent is married.
employ	In Paid Employment	=1 if in paid employment. Changes made, cf. Appendix C.2.
foreign	Foreign-born	=1 if the respondent declares being born abroad. In this analysis, immigrants are all those who were born abroad.
for04	Post-04 Immigrant	=1 if the respondent if foreign-born and declares having arrived in – or after – 2004.
regionborn	Macro-region of provenance	The area of the world of provenance, derived from the country of origin. 0 is reserved for those born in the UK.
exp	Potential Experience	Years elapsed since reported age at completion of education. exp^2 is the relative square.
forexp	Potential Experience if Immigrant	=0 if UK-born, =exp if Foreign-born. $forexp^2$ is the relative square.

Continued on next page

Table A1 – *Continued from previous page*

First entry	Second entry	Third entry
ukexp	Potential British Experience	Years elapsed since arrival to the UK (if immigrant); Years elapsed since completion of schooling (if the respondent migrated before completing education). $ukexp^2$ is the relative square.
homexp	Potential Home Experience	Years elapsed between leaving education and emigrating to the UK (if immigrant). $homexp^2$ is the relative square.
expInt	Experience Interaction	Interaction of Home and British Experience. It serves to identify possible effects of combining the two beyond addition.
ysa	Years Since Arrival	Years elapsed since declared arrival to the UK. YSA^2 is the relative square.
educ	Highest Qualification	Highest educational qualification attained.
wage	Net Labour Income	Labour component of total personal income, net of taxes.
gor_dv	UK Region of residence	Derived region of residence of the interviewee.
jbsic_07	Current Job	Job reported by the respondent, coded following the SIC 2007 1-digit division.
dec#	Decade of Arrival Dummy	Reported decade of arrival (if immigrant), derived from year of arrival. Period of arrival is a standard proxy for immigrant cohort.

C.2 $employ_{it}$ Determination

There are two main questions that refer to employment in the Main Survey (*indresp*) datafile, which contains answers from the individual interview for each wave: *employ* and *jbstat*[31]. *Employ* records the answer to the question "Are you in paid employment?", asked to every individual aged 16 or higher. Excluding proxy answers, there are four answers available: "yes", "no", "don't know" and refusal to respond[32]. *Jbstat* instead presents individuals with a wider palette of responses to the question "Which of these best describes your current employment situation?": these include, among many, "self-employed", "unemployed", "retired", "full-time student" and the broader "doing something else"[33]. Contrary to what I expected, there is a significant amount of individuals for which answers to the two questions diverge: for example, in the raw downloadable data for Wave 1 there are 3271 individuals that answer "yes" to *employ* and "unemployed" to *jbstat*. Such a high number of incoherences hints to more than random noise.

Another problem is represented by the fact that reported net labour income *inc1lab*, one of the two measures of wage I make use of (the other being gross income *fimnlabgrs_dv*, which is anyway derived post-field and as such less preferable – regression results are similar independently of the choice), is positive for some individuals who answer *employ* negatively. Referring again to Wave 1, 282 respondents report net labour income in excess of £800 while not being in paid employment; 48 report earning similar amounts while responding "unemployed" to *jbstat*.

Dropping problematic observations, combined with the willingness to have a balanced panel, would have yielded a greatly reduced sample size as the entire individual history

[31] Variable names in this subsection are taken directly from the STATA files from the EUL download.

[32] As expected, responses are for the vast majority either positive or negative: tellingly, in Wave 1 only 8 out of 50994 responses is something other than "yes" or "no".

[33] Please refer to the questionnaire PDFs available at https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires for information.

would have been eliminated. To preserve the sample, arbitrary decisions in recoding the $employ_{it}$ dichotomous variable were made, making use of responses to *jbstat*, reported earnings and the National Minimum Wage. Starting from the raw *employ* answers, I first code as employed individuals that report monthly earnings higher than a threshold dependent on the year- and age-specific minimum wage, while defining those who earn below the threshold unemployed. Such threshold is calculated as

$$thr_{it} = \frac{minwage_{it} * 16 * 52}{12} \tag{16}$$

where $minwage_{it}$ – hourly minimum wage – varies by year and age of respondent[34] and 16 is the maximum number of hours one can work while receiving Jobseekers' Allowance – the reception of which is recorded. This redefinition of employment based on thr_{it} interests a low number of individuals, supporting the idea that employment and income do not contradict each other consistently but rather as result of noise.

The next step consists in making use of information on benefits received: respondents of working age are asked whether they receive unemployment benefits, income support, incapacity pension, severe disablement allowance, employment allowance and other state benefits related to unemployment[35]. I make use of receipt to determine unemployment status.

Last, I make use of responses to *jbstat*: in particular, this allows me to set as employed women on maternity leave (who could answer negatively to *employ* but still receive labour income), self-employed (who could earn less than minimum wage) and part-time workers (in the unlikely case that some of them weekly earn less than the equivalent of 16 hours of minimum wage) who were set as unemployed by using thr_{it}.

[34] There are in the United Kingdom two different minimum wages,a lower one for workers younger than 21 and a higher one for older workers

[35] Question names and specifics are reported in the question cards available online at https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires, Wave 3 Main questionnaire. The subselection of the questions used is also available from the author as additional appendix.

D Auxiliary Tests and Robustness Checks

D.1 Returns to exp_{it} for natives and immigrants: Earnings

Table A2: FGLS with Mundlak (1978) Correction Estimates of (1) and (2) – including $forexp_{it}$

	(1)	(2)	(3)	(4)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap
Male	0.408***	0.414***	0.408***	0.414***
	(0.01)	(0.01)	(0.01)	(0.01)
Immigrant	-0.294**		-0.186	
	(0.15)		(0.17)	
Post-04 Immigrant			-0.095	-0.080
			(0.06)	(0.06)
Potential Experience	0.040***	0.041***	0.040***	0.041***
	(0.00)	(0.00)	(0.00)	(0.00)
Potential Experience Squared x0.01	-0.037***	-0.038***	-0.037***	-0.039***
	(0.01)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK	0.015	0.029***	0.019	0.027***
	(0.02)	(0.01)	(0.02)	(0.01)
Years Since Arrival to the UK Squared x0.01	-0.049***	-0.052***	-0.049***	-0.048***
	(0.02)	(0.02)	(0.02)	(0.02)
$M_i * exp_{it}$	0.016	-0.003	0.012	-0.003
	(0.02)	(0.01)	(0.02)	(0.01)
$M_i * exp_{it}^2$ x0.01	-0.017	-0.007	-0.017	-0.008
	(0.03)	(0.02)	(0.03)	(0.02)
European		-0.231**		-0.198**
		(0.10)		(0.10)
Asian		-0.517***		-0.491***
		(0.10)		(0.10)
African		-0.364***		-0.336***
		(0.10)		(0.10)
North American		-0.241*		-0.219
		(0.13)		(0.13)
Centre-South American		-0.386***		-0.361***
		(0.12)		(0.12)
Australian or New Zealander		0.010		0.035
		(0.13)		(0.14)
Non-specified Countries		-0.382***		-0.357***
		(0.10)		(0.10)
No Qualification		*Term of Comparison*		
Other Qualification	0.127***	0.114***	0.127***	0.115***
	(0.03)	(0.03)	(0.03)	(0.03)
GCSE	0.121***	0.113***	0.121***	0.114***
	(0.03)	(0.03)	(0.03)	(0.03)
A Levels	0.242***	0.231***	0.242***	0.231***
	(0.03)	(0.03)	(0.03)	(0.03)
Degree	0.620***	0.610***	0.620***	0.610***
	(0.03)	(0.03)	(0.03)	(0.03)
Observations	47,333	47,333	47,333	47,333
Number of individuals	11,190	11,190	11,190	11,190

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 .
Omitted output: Marriage, Wave Fixed Effects, Decade of Arrival FE, Job FE, Year*Region FE, Mundlak Correction

Employing the terminology of equation (3), one can expect β_1^i to be equal to the sum of the coefficient attached to exp_{it} and the coefficient of $M_i * exp_{it}$, while $\beta_1^n = exp_{it}$. In fact, including $M_i * exp_{it}$ in equation (1) – which allows to drop the i and n superscripts – yields (leaving YSA_{it} aside)

$$\frac{\partial ln(wage)_{it}}{\partial t}|_{immigrant} = \beta_1 + 2\beta_2 exp_{it} + \beta_{forexp} * 1 + 2\beta_{forexp2} * 1 * exp_{it} \tag{17}$$

$$\frac{\partial ln(wage)_{it}}{\partial t}|_{native} = \beta_1 + 2\beta_2 exp_{it} + \beta_{forexp} * 0 + 2\beta_{forexp2} * 0 * exp_{it} \tag{18}$$

and

$$\frac{\partial ln(wage)_{it}}{\partial t}|_{immigrant} - \frac{\partial ln(wage)_{it}}{\partial t}|_{native} = \beta_{forexp} * 1 * +2\beta_{forexp2} * 1 * exp_{it} \tag{19}$$

The fact that $M * exp_{it}$ is insignificant in table A2 highlights the fact that, in this sample, there is no evidence of lower returns to experience among immigrants.
It then follows that equation (3) can be reduced to equation (4).

D.2 WG and POLS Estimates of equations (1) and (2)

Table A3: Pooled OLS Estimates of equations (1) and (2)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-0.407***		-0.362***		-0.235***	
	(0.07)		(0.08)		(0.08)	
Post-04 Immigrant			-0.049	-0.078**	-0.048	-0.077**
			(0.03)	(0.03)	(0.03)	(0.03)
Potential Experience	0.025***	0.025***	0.025***	0.025***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.049***	-0.049***	-0.049***	-0.049***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential UK Experience					0.026***	0.026***
					(0.00)	(0.00)
Potential UK Experience Squared x0.01					-0.050***	-0.050***
					(0.00)	(0.00)
Potential Home Experience (Immigrants only)					0.012*	0.016**
					(0.01)	(0.01)
Potential Home Experience Squared x0.01 (Immigrants only)					-0.030*	-0.043**
					(0.02)	(0.02)
Home and UK Exp Interaction (Immigrants only)					-0.046*	-0.051*
					(0.03)	(0.03)
Years Since Arrival to the UK	0.023***	0.025***	0.019***	0.018**	0.006	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK Squared x0.01	-0.044***	-0.047***	-0.037***	-0.037***	-0.006	-0.007
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
European		-0.267***		-0.191**		-0.094
		(0.07)		(0.08)		(0.08)
Asian		-0.549***		-0.478***		-0.371***
		(0.07)		(0.08)		(0.08)
African		-0.406***		-0.333***		-0.227***
		(0.08)		(0.08)		(0.08)
North American		-0.292***		-0.227**		-0.116
		(0.09)		(0.09)		(0.09)
Centre-South American		-0.463***		-0.394***		-0.293***
		(0.09)		(0.09)		(0.09)
Australian or New Zealander		-0.048		0.021		0.141
		(0.08)		(0.09)		(0.09)
Non-specified Country		-0.394***		-0.325***		-0.221***
		(0.08)		(0.08)		(0.09)
Observations	47,333	47,333	47,333	47,333	47,333	47,333
R^2	0.246	0.249	0.246	0.249	0.246	0.249

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Omitted output: Male dummy, Education, Marriage Dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE

Table A4: Within Group Estimates of equations (1) and (2)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-		-		-	
Post-04 Immigrant			-	-	-	-
Potential Experience	0.041***	0.041***	0.041***	0.041***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.039***	-0.039***	-0.039***	-0.039***		
	(0.01)	(0.01)	(0.01)	(0.01)		
Potential UK Experience					0.041***	0.041***
					(0.00)	(0.00)
Potential UK Experience Squared x0.01					-0.039***	-0.039***
					(0.01)	(0.01)
Potential Home Experience (Immigrants only)					-	-
Potential Home Experience Squared x0.01 (Immigrants only)					-	-
Home and UK Exp Interaction (Immigrants only)					-0.058	-0.058
					(0.07)	(0.07)
Years Since Arrival to the UK	0.027***	0.027***	0.027***	0.027***	0.023**	0.023**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK Squared	-0.052***	-0.052***	-0.052***	-0.052***	-0.035**	-0.035**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
European		-		-		-
Asian		-		-		-
African		-		-		-
North American		-		-		-
Centre-South American		-		-		-
Australian or New Zealander		-		-		-
Non-specified Country		-		-		-
Observations	47,333	47,333	47,333	47,333	47,333	47,333
R^2	0.020	0.020	0.020	0.020	0.019	0.019
Number of individuals	11,190	11,190	11,190	11,190	11,190	11,190

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Omitted output: Male dummy, Education, Marriage Dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE.

D.3 Logit and Conditional FE Logit AMEs of equations (9) and (10)

Table A5: AMEs for the Pooled Logit ML Estimator of equations (9) and (10)

VARIABLES	dy/dx	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
Male	.106	.003	34.27	0.000	.100	.113
Potential UK Experience	.017	.001	31.18	0.000	.016	.018
Potential UK Experience Squared x0.01	-.039	.001	-35.63	0.000	-.041	-.037
Years Since Arrival	-.009	.001	-6.21	0.000	-.012	-.006
Years Since Arrival Squared x0.01	.026	.003	10.24	0.000	.021	.031
Potential Home Experience (Immigrants only)	.013	.002	6.14	0.000	.009	.017
Potential Home Experience Squared x0.01 (Immigrants only)	-.022	.006	-3.70	0.000	-.033	-.010
Home and UK Exp Interaction (Immigrants only)	-.073	.008	-8.92	0.000	-.090	-.057
Immigrant	-.036	.019	-1.89	0.058	-.073	.001
Post-04 Immigrant	.009	.013	0.69	0.491	-.016	.034
No Qualification	*term of comparison*					
Other Qualification	.121	.006	20.41	0.000	.109	.132
GCSE	.177	.005	34.65	0.000	.167	.187
A Levels	.251	.005	47.44	0.000	.241	.261
Degree	.337	.005	63.90	0.000	.327	.347
N	70,940	70,940	70,940	70,940	70,940	70,940

Notes: dy/dx for factor levels is the discrete change from the base level.

Omitted output: Wave Fixed Effects. Region FE, Region of birth FE and Year*Region FE not included.

Table A6: AMEs for the Conditional FE Logit ML Estimator of equations (9) and (10)

VARIABLES	dy/dx	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
Male	-					
Potential UK Experience	.016	.007	2.36	0.018	.003	.029
Potential UK Experience Squared x0.01	-.061	.0123	-5.00	0.000	-.086	-.037
Years Since Arrival	.007	.014	0.48	0.628	-.021	.035
Years Since Arrival Squared x0.01	.003	.026	0.13	0.896	-.049	.054
Potential Home Experience (Immigrants only)	-					
Potential Home Experience Squared x0.01 (Immigrants only)	-					
Home and UK Exp Interaction (Immigrants only)	-.072	.091	-0.80	0.424	-.250	.105
Immigrant	-					
Post-04 Immigrant	-					
No Qualification	*term of comparison*					
Other Qualification	-					
GCSE	-					
A Levels	-					
Degree	-					
N	13,510	13,510	13,510	13,510	13,510	13,510

Notes: Conditional Fixed Effects considers only individuals with within-variation in the dependent variable.

Omitted output: Wave Fixed Effects. Region FE, Region of birth FE and Year*Region FE not included.

D.4 Including weights in POLS and WG estimation

Table A7: Weighted WG Estimates of equations (1) and (2) – Selected Output

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-		-		-	
Post-04 Immigrant			-	-	-	-
Potential Experience	0.043***	0.043***	0.043***	0.043***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.045***	-0.045***	-0.045***	-0.045***		
	(0.01)	(0.01)	(0.01)	(0.01)		
Potential UK Experience					0.041***	0.041***
					(0.00)	(0.00)
Potential UK Experience Squared x0.01					-0.039***	-0.039***
					(0.01)	(0.01)
Potential Home Experience (Immigrants only)					-	-
Potential Home Experience Squared x0.01 (Immigrants only)					-	-
Home and UK Exp Interaction (Immigrants only)					-0.058	-0.058
					(0.07)	(0.07)
Years Since Arrival to the UK	0.013	0.013	0.013	0.013	0.005	0.005
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Years Since Arrival to the UK Squared	-0.045	-0.045	-0.045	-0.045	-0.020	-0.020
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
European		-		-		-
Asian		-		-		-
African		-		-		-
North American		-		-		-
Centre-South American		-		-		-
Australian or New Zealander		-		-		-
Non-specified Country		-		-		-
Observations	46,890	46,890	46,890	46,890	46,890	46,890
R-squared	0.020	0.020	0.020	0.020	0.020	0.020
Number of pidp	11,086	11,086	11,086	11,086	11,086	11,086

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 .

Omitted Output: Male Dummy, Education, Marriage Dummy, Decade FE, Job FE, Region*Year FE, Wave FE

Table A8: Weighted POLS Estimates of equations (1) and (2) – Selected Output

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-0.342***		-0.256**		-0.139	
	(0.11)		(0.12)		(0.13)	
Post-04 Immigrant			-0.096**	-0.105***	-0.104**	-0.113***
			(0.04)	(0.04)	(0.04)	(0.04)
Potential Experience	0.028***	0.028***	0.028***	0.028***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.056***	-0.056***	-0.056***	-0.057***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential UK Experience					0.026***	0.026***
					(0.00)	(0.00)
Potential UK Experience Squared x0.01					-0.050***	-0.050***
					(0.00)	(0.00)
Potential Home Experience (Immigrants only)					0.012*	0.016**
					(0.01)	(0.01)
Potential Home Experience Squared x0.01 (Immigrants only)					-0.030*	-0.043**
					(0.02)	(0.02)
Home and UK Exp Interaction (Immigrants only)					-0.046*	-0.051*
					(0.03)	(0.03)
Years Since Arrival to the UK	0.026**	0.026**	0.018*	0.018*	0.007	0.007
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK Squared x0.01	-0.050***	-0.052***	-0.038*	-0.039*	-0.010	-0.011
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
European		-0.258**		-0.159		-0.082
		(0.10)		(0.11)		(0.12)
Asian		-0.485***		-0.393***		-0.296**
		(0.11)		(0.11)		(0.13)
African		-0.329***		-0.233**		-0.145
		(0.11)		(0.11)		(0.13)
North American		-0.189		-0.111		-0.017
		(0.12)		(0.12)		(0.14)
Centre-South American		-0.432***		-0.341***		-0.258*
		(0.12)		(0.12)		(0.14)
Australian or New Zealander		-0.033		0.058		0.168
		(0.12)		(0.12)		(0.14)
Non-specified Country		-0.358***		-0.265**		-0.178
		(0.13)		(0.13)		(0.15)
Observations	46,890	46,890	46,890	46,890	46,890	46,890
R^2	0.251	0.252	0.251	0.253	0.250	0.252

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Omitted output: Male Dummy, Education, Marriage Dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE

D.5 Testing correlation of θ_i and x_{it}

As detailed in Section 4.1, the Feasible GLS estimator is inconsistent in presence of correlation between θ_i and x_{it}: if such correlation exists, one can either use the WG estimator – which will be consistent although inefficient – or model the correlation by employing either the Mundlak (1978) or Chamberlain (1980) specification. If however such correlation does not exist, Feasible GLS will be efficient and thus preferrable over other methods.

We can test for the presence of correlation between covariates and individual heterogeneity in two main ways: by testing the significance of the Mundlak (1978) correction – i.e. testing whether coefficients of within averages of time-dependent covariates are significantly different from zero –, as already suggested in the original Mundlak paper and in Wooldridge (2010); or by Hausman (1978) test, which tests whether $\hat{q} = \hat{\beta}_{WG} - \hat{\beta}_{FGLS}$ is equal to zero, as it should be under no misspecification.

Below are the results for the tests: all three reject respective null hypotheses, pointing to significant correlation of individual heterogeneity with the covariates.

Hausman Test, using equation (2) specification

β_{WG} = consistent under H_0 and H_1; inefficient under H_0

β_{FGLS} = inconsistent under H_1; efficient under H_0

H_0: Difference in coefficients not systematic

H_1: H_0 is false

$\chi^2_{112} = (\hat{\beta}_{WG} - \hat{\beta}_{FGLS})'[\hat{V}_{WG} - \hat{V}_{FGLS}]^{-1}(\hat{\beta}_{WG} - \hat{\beta}_{FGLS}) = 441.68$

$Prob > \chi^2 = 0.0000$

Robust Hausman Test, using *xtoverid*

Test of Overidentifying Restrictions: Within Group vs Feasible GLS

Cross-section Time-series model: xtreg re robust cluster(pidp)

Sargan-Hansen Statistic: 400.427

χ^2_{112} P-value = 0.0000

Testing significance of Mundlak correction

This method implies estimating an augmented version of equation (2), storing related coefficient estimates and testing the joint significance of the Mundlak corrections.

Table A9: FGLS estimation of equation (2) - Selected Output

VARIABLES	
Mean($ukexp_{it}$)	-0.012***
	(0.00)
Mean($ukexp^2_{it}$)	-0.017*
	(0.01)
Mean($ukexp_{it} * homexp_i$)	0.038
	(0.08)
Mean(YSA_{it})	-0.029
	(0.02)
Mean(YSA^2_{it})	0.047
	(0.04)
Observations	47,333
Number of pidp	11,190

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

(1) Mean($ukexp_{it}$) = 0

(2) Mean($ukexp^2_{it}$) = 0

(3) Mean($ukexp_{it} * homexp_i$) = 0

(4) Mean(YSA_{it}) = 0

(5) Mean(YSA^2_{it}) = 0

$\chi^2_5 = 137.71$

$Prob > \chi^2 = 0.0000$

D.6 Testing for AR(1) error term and Baltagi Wu (1999) estimates

The test for error term autocorrelation presented in Wooldridge (2010) hinges on the assumption of homoskedasticity (and thus no serial correlation) of the error term, which ensures efficiency of the Within estimator. If the error term ϵ_{it} of the specified equation is uncorrelated, the author observes that the error term $e_{it} \equiv \Delta\epsilon_{it}$ of the first-differenced equation will present first-order autocorrelation of -0.5. The process then consists in storing the residuals of a regression in first-differences and, by testing whether e_{it} presents autocorrelation of -0.5, obtain information on the structure of the principal equation.

The test is performed in STATA through the user-written command *xtserial*, presented in Drukker (2003). I use equation (1) for the test which, making use of first-differences, eliminates all time-invariant covariates. The hypothesis of serial autocorrelation is rejected.

Wooldridge Test for Autocorrelation in Panel Data

H_0: No first order autocorrelation

$F(1, 8876) = 0.162$

$Prob > F = 0.6874$

Baltagi Wu (1999) Estimates

If such AR(1) correlation was present in the data, one could use the Baltagi Wu (1999) estimator (specifically designed for unbalanced panels): first, such estimator implies estimating the autocorrelation coefficient; it then corrects the data by netting out the effect of autocorrelation, and finally estimates the specified equation by FGLS or WG using the computed data. Applying such estimator to the data yields

Table A10: Baltagi Wu (1999) estimation of equations (1) and (2)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Gap at entry	Gap at entry	Post-04 Gap	Post-04 Gap	Dividing Experience	Dividing Experience
Immigrant	-0.384***		-0.285**		-0.113	
	(0.11)		(0.13)		(0.14)	
Post-04 Immigrant			-0.089	-0.082	-0.090	-0.126**
			(0.06)	(0.05)	(0.06)	(0.06)
Potential Experience	0.041***	0.041***	0.041***	0.041***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Potential Experience Squared x0.01	-0.040***	-0.040***	-0.040***	-0.041***		
	(0.01)	(0.01)	(0.01)	(0.01)		
Potential UK Experience					0.040***	0.040***
					(0.00)	(0.00)
Potential UK Experience Squared x0.01					-0.038***	-0.038***
					(0.01)	(0.01)
Potential Home Experience (Immigrants only)					0.009	0.013
					(0.01)	(0.01)
Potential Home Experience Squared x0.01 (Immigrants only)					-0.028	-0.042
					(0.03)	(0.03)
Home and UK Exp Interaction (Immigrants only)					-0.079	-0.077
					(0.07)	(0.07)
Years Since Arrival to the UK	0.028***	0.026***	0.028***	0.023***	0.027***	0.027***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Years Since Arrival to the UK Squared x0.01	-0.056***	-0.054***	-0.056***	-0.049***	-0.042**	-0.043**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
European		-0.265***		-0.225***		0.070
		(0.08)		(0.08)		(0.14)
Asian		-0.561***		-0.528***		-0.225*
		(0.07)		(0.08)		(0.14)
African		-0.411***		-0.375***		-0.071
		(0.08)		(0.08)		(0.14)
North American		-0.287***		-0.258**		0.046
		(0.11)		(0.11)		(0.15)
Centre-South American		-0.451***		-0.418***		-0.123
		(0.11)		(0.11)		(0.16)
Australian or New Zealander		-0.030		0.003		0.318*
		(0.12)		(0.12)		(0.17)
Non-specified Country		-0.421***		-0.388***		-0.090
		(0.07)		(0.08)		(0.14)
ρ			.11698806			
Observations	47,333	47,333	47,333	47,333	47,333	47,333
Number of pidp	11,190	11,190	11,190	11,190	11,190	11,190

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 .

Omitted output: Male dummy, Education, Marriage dummy, Year FE, Wave FE, Job FE, Decade of Arrival FE, Region of residence FE, Residence*Year FE, Mundlak correction

Coefficients present no major variation from those obtained in Table 4, and the estimated ρ is very low – consistently with the Wooldridge test.

References

Arellano, Manuel and Stephen Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *The Review of Economic Studies* 58.2, pp. 277–297.

Baltagi, Badi and Ping X. Wu (1999). "Unequally Spaced Panel Data Regressions with AR(1) Disturbances". In: *Econometric Theory* 15.06, pp. 814–823.

Becker, Gary S. (1964). *Human Capital*. Columbia University Press.

Beecham, Lord Jeremy (2016a). *Social Security Benefits and Taxation: EU Nationals: Written question – HL183*. URL: http://goo.gl/b7x9LE (visited on 09/03/2016).

— (2016b). *Social Security Benefits and Taxation: EU Nationals: Written question – HL5912*. URL: http://goo.gl/WgvDnX (visited on 09/03/2016).

Beenstock, Michael, Barry R. Chiswick, and Ari Paltiel (2010). "Testing the Immigrant Assimilation Hypothesis with Longitudinal Data". In: *Review of Economics of the Household* 8.1, pp. 7–27.

Bell, Brian D. (1997). "The Performance of Immigrants in the United Kingdom: Evidence from the GHS". In: *The Economic Journal* 107.441, pp. 333–344.

Blackaby, D. H. et al. (1994). "Black-white Male Earnings and Employment Prospects in the 1970s and 1980s: Evidence for Britain". In: *Economics Letters* 46.3, pp. 273–279.

Blackaby, David H (1986). "An Analysis of the Male Racial Earnings Differential in the UK, using the General Household Survey". In: *Applied Economics* 18.11, pp. 1233–1242.

Borjas, George J. (1985). "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants". In: *Journal of Labor Economics* 3.4, pp. 463–89.

Borjas, George J. (1989). “Immigrant and Emigrant Earnings: A Longitudinal Study”. In: *Economic Inquiry* 27.1, pp. 21–37.

— (1994). “The Economics of Immigration”. In: *Journal of Economic Literature* 32.4, pp. 1667–1717.

— (1999). “The Economic Analysis of Immigration”. In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. 1st ed. Vol. 3, Part A. North Holland. Chap. 28, pp. 1697–1760.

— (2013). “The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again”. In: *National Bureau of Economic Research Working Paper Series 19116*.

Butler, John S. and Robert Moffitt (1982). “A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model”. In: *Econometrica* 50.3, pp. 761–764.

Cadena, Brian C., Brian Duncan, and Stephen J. Trejo (2015). “The Labor Market Integration and Impacts of US Immigrants”. In: *Handbook of the Economics of International Migration*. Ed. by Barry R. Chiswick and Paul W. Miller. Vol. 1. Handbook of the Economics of International Migration. North-Holland, pp. 1197 –1259.

Cameron, Colin and Pravin Trivedi (2005). *Microeconometrics*. Cambridge University Press.

— (2010). *Microeconometrics Using Stata*. StataPress, Revised Edition.

Chamberlain, Gary (1980). “Analysis of Covariance with Qualitative Data”. In: *The Review of Economic Studies* 47.1, pp. 225–238.

— (1984). “Panel Data”. In: *Handbook of Econometrics*. Ed. by Z. Griliches and M. D. Intriligator. 1st ed. Vol. 2. North-Holland. Chap. 22, pp. 1247–1318.

Chetty, Raj (2012). *Time Trends in the Use of Administrative Data for Empirical Research*. URL: http://www.rajchetty.com/chettyfiles/admin_data_trends.pdf (visited on 09/07/2016).

Chiswick, Barry R. (1978). "The Effect of Americanization on the Earnings of Foreign-born Men". In: *Journal of Political Economy* 86.5, pp. 897–921.

— (1980). "The Earnings of White and Coloured Male Immigrants in Britain". In: *Economica* 47.185, pp. 81–87.

Chiswick, Barry R., Yew Liang Lee, and Paul W. Miller (2005). "Immigrant Earnings: a Longitudinal Analysis". In: *Review of Income and Wealth* 51.4, pp. 485–503.

Clark, Ken and Stephen Drinkwater (2008). "The Labour Market Performance of Recent Migrants". In: *Oxford Review of Economic Policy* 24.3, pp. 495–516.

Clark, Ken and Jo Lindley (2009). "Immigrant assimilation pre and post labour market entry: evidence from the UK Labour Force Survey". In: *Journal of Population Economics* 22.1, pp. 175–198.

Dickens, Richard and Abigail McKnight (2008). "Assimilation of Migrants into the British Labour Market". In: *LSE STICERD Research Paper No. CASE133*.

Drukker, David M. (2003). "Testing for serial correlation in linear panel-data models". In: *Stata Journal* 3.2, pp. 168–177.

Duleep, Harriet Orcutt and Mark C Regets (2002). "The Elusive Concept of Immigrant Quality: Evidence from 1970-1990". In: *IZA Discussion Paper No. 631*.

Dustmann, Christian, Francesca Fabbri, and Ian Preston (2005). "The Impact of Immigration on the British Labour Market". In: *The Economic Journal* 115.507, F324–F341.

Dustmann, Christian and Tommaso Frattini (2014). "The Fiscal Effects of Immigration to the UK". In: *The Economic Journal* 124.580, F593–F643.

Dustmann, Christian, Tommaso Frattini, and Ian P Preston (2013). “The Effect of Immigration Along the Distribution of Wages”. In: *The Review of Economic Studies* 80.1, pp. 145–173.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt (1998). “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics”. In: *National Bureau of Economic Research Working Paper 220*.

Friedberg, Rachel M. (2000). “You Can’t Take It with You? Immigrant Assimilation and the Portability of Human Capital”. In: *Journal of Labor Economics* 18.2, pp. 221–251.

Funkhouser, Edward and Stephen J. Trejo (1995). “The Labor Market Skills of Recent Male Immigrants: Evidence from the Current Population Survey”. In: *Industrial & Labor Relations Review* 48.4, pp. 792–811.

Greene, William (2004). “The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the presence of Fixed Effects”. In: *Econometrics Journal* 7.1, pp. 98–119.

Greene, William H. (2012). *Econometric Analysis*. Pearson, Seventh Edition.

Hausman, Jerry A. (1978). “Specification Tests in Econometrics”. In: *Econometrica*, pp. 1251–1271.

HM Government (2014). *Review of the Balance of Competences between the United Kingdom and the European Union - EU Enlargement*. URL: http://goo.gl/uZtcMo (visited on 09/03/2016).

Hsiao, Cheng (2014). *Analysis of Panel Data*. Cambridge University Press, Third Edition.

Knies, Gundi (2015). *Understanding Society – UK Household Longitudinal Study: Wave 1-5, 2009-2014, User Manual*. University of Essex, Institute for Social and Economic Research.

LaLonde, Robert J. and Robert H Topel (1992). "The Assimilation of Immigrants in the US Labor Market". In: *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*. The University of Chicago Press, pp. 67–92.

Lemos, Sara (2013). "Immigrant Economic Assimilation: Evidence from UK Longitudinal Data between 1978 and 2006". In: *Labour Economics* 24, pp. 339 –353.

Licht, Georg and Viktor Steiner (1994). "Assimilation, Labour Market Experience and Earnings Profiles of Temporary and Permanent Immigrant Workers in Germany". In: *International Review of Applied Economics* 8.2, pp. 130–156.

Longhi, Simonetta and Magdalena Rokicka. "European immigrants in the UK before and after the 2004 Enlargement: Is there a change in Immigrant self-selection?" In: *CreAM Discussion Paper Series n. 32/12*.

Lubotsky, Darren (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings". In: *Journal of Political Economy* 115.5, pp. 820–867.

Lynn, Peter et al. (2012). "An Initial Look at Non-Response and Attrition in Understanding Society". In: *Understanding Society Working Paper Series: 2012-02*.

Manacorda, Marco, Alan Manning, and Jonathan Wadsworth (2012). "The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain". In: *Journal of the European Economic Association* 10.1, pp. 120–151.

Migration Advisory Committee (2014). *Migrants in Low-Skilled Work*. Tech. rep. Home Office.

Mundlak, Yair (1978). "On the Pooling of Time Series and Cross Section Data". In: *Econometrica*, pp. 69–85.

Office for National Statistics (2014). *2011 Census analysis: Ethnicity and the Labour Market, England and Wales*. URL: http://goo.gl/oYRFJ5 (visited on 09/04/2016).

— (2015). *Participation rates in the UK – 2014 – 2. Women*. URL: http://goo.gl/cSkb50 (visited on 09/04/2016).

Pischke, Jörn-Steffen (1992). "Assimilation and the Earnings of Guestworkers in Germany". In: *Zentrum für Europäische Wirtschaftsforschung Mannheim No. 92-17.*

Stewart, Mark B. (1983). "Racial Discrimination and Occupational Attainment in Britain". In: *The Economic Journal* 93.371, pp. 521–541.

University of Essex, Institute for Social and Economic Research, NatCen Social Research. (2015). *Understanding Society: Waves 1-5, 2009-2014. [data collection]. 7th Edition.* UK Data Service. SN: 6614.

Wheatley Price, Stephen (2001). "The Employment Adjustment of Male Immigrants in England". In: *Journal of Population Economics* 14.1, pp. 193–220.

Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Second Edition.