Immigrant Economic Assimilation:

Evidence from UK Longitudinal Data between 1978 and 2006

Sara Lemos, *University of Leicester*[1]

October 2013

Using the underexplored, sizeable and long Lifetime Labour Market Database (LLMDB) we estimate the immigrant-native earnings gap at entry and over time for the UK between 1978 and 2006. That is, we attempt to separately estimate cohort and assimilation effects. We also estimate the associated immigrant earnings growth rate and immigrant-native earnings convergence rate. Our estimates suggest that immigrants from more recent cohorts fare better than earlier ones at entry. Furthermore, the earnings of immigrants from more recent cohorts catch up faster with natives' earnings. While the convergence took over 30 years for those entering in the post-war, it only took half as long for those entering in the early 2000s. This earnings growth is fastest in the first 10 years, and it considerably slows down after 30 years.

JEL classification: J24 J31 J61 J71 J82 F22.

Keywords: immigration, assimilation, wages, earnings, earnings-gap, UK.

1. Introduction

Immigrant assimilation has been, and continues to be, on the forefront of immigration literature for the past 30 years. Much of this literature is concerned with the labour market adjustment of immigrants in the host country. In particular, this literature has focused on estimating the immigrant-native earnings gap at entry and over time, as well as the associated immigrant earnings growth rate and the immigrant-native earnings convergence rate, which is sometimes referred to as immigrant (earnings) assimilation.

The early cross-section studies in this literature found that immigrants' earnings were lower than the earnings of comparable natives at entry. However, as immigrants' earnings increased at a faster rate – fastest in the first few years after immigration – the initial immigrant-native earnings gap closed over time. This earnings convergence, it was argued, derived from the fact that immigrants were more able, more motivated or more hardworking than natives.

There was an intense debate in the literature about the interpretation of these findings. The main criticism was that, instead of providing a measure of earnings assimilation, the cross-section evidence was actually capturing changes over time in unmeasured dimensions of immigrants' skills, such as a decline in the relative skills of more recent cohorts (Chiswick 1980; Borjas 1985). In that case, the cross-section evidence had little to say about immigrant-native earnings convergence. Other such selection arguments included return migration of immigrants that were less able and other types of data attrition.

Biases implied by these various types of selectivity in the data, such as cohort bias and survivor bias, became a central identification issue in the literature (Chiswick 1978 and 1980; Carliner 1980; Borjas 1985, 1994 and 1999; Duleep and Regets 1997; Chiswick et al. 2005). This literature soon recognized that identification of assimilation effects separately from cohort effects required a sufficiently large and long longitudinal dataset that tracked particular workers over time – a very scarce kind of dataset.

The main contribution of this paper is to exploit one such dataset to estimate assimilation effects separately from cohort effects. We use the UK Lifetime Labour Market Database (LLMDB), which is a large and long longitudinal dataset – yet largely unused in the immigration literature – that combines anonymised administrative tax records and social security records into a dataset that tracks a random sample of 647,000 individuals between 1978 and 2006. The LLMDB not only tracks workers over their working lives, but also has high levels of accuracy and relatively low levels of attrition. Its longitudinal nature combined with its large sample size and long sample period enables us to achieve our objective of

separately estimating assimilation and cohorts effects. Given that the availability of longitudinal data has been so limited, this paper is an important contribution to the literature.

Furthermore, this paper is an important contribution to the rather limited UK immigration literature, where assimilation and cohort effects estimates are largely unavailable.[2] This is in contrast with the international literature, where considerable effort has been put into estimating such effects. In his pioneering paper, Chiswick (1980) remarked on such scarcity in the UK literature. Nevertheless, 30 years on, less than a handful of papers have followed (Bell 1997; Dustmann and Fabbri 2005; Dickens and McKnight 2008). Therefore, this paper is an important contribution as it helps to fill this blank in the UK literature and it helps to inform policymaking in the face of continuing public debate on immigration policy in the UK.

We initially utilize our longitudinal data as a succession of independent cross-sections to estimate a simple descriptive model of the evolution of the earnings gap between immigrants and natives over time. We then re-estimate our simple model using a restricted sample of immigrants who stayed in the UK for 30 years or longer to account for survivor bias. Finally, we exploit the longitudinal nature of our data to estimate a more complete earnings gap model in an attempt to disentangle cohort and assimilation effects. Our approach confirms – in line with other results in the literature – that cross-section data estimates are biased and a comparison of these with longitudinal data estimates gives an indication of the extent of this bias – which is a contribution to the literature.

Our results suggest that immigrants from more recent cohorts fare better than earlier ones at entry. For example, whereas immigrants that arrived in the post-war period earned nearly 60% less than comparable natives at entry, those arriving in the early 2000s earned only 30% less at entry. Furthermore, our estimates suggest that the earnings of immigrants from more recent cohorts catch up faster with natives' earnings. While the convergence took over 30 years for those entering in the post-war, it only took half as long for those entering in the early 2000s. The earnings of immigrants increase 33% faster than the earnings of natives in the first 10 years after arrival. This earnings convergence is fastest in the first 10 years, and it

---

[2] Although the corresponding literature is quite large for the US (Chiswick 1978; Borjas 1985, 1995 and 1999; Butcher and DiNardo 2002; Cortes 2004; Chiswick et al. 2005; etc.), it is limited for the UK. Using data from the 1972 General Household Survey (GHS) to estimate a standard human capital earnings model, Chiswick (1980) found no earnings gap for white but a -25% gap for non-white male immigrants. In an attempt to model cohort and assimilation effects separately, Bell (1997) used 1973-1992 GHS data and broadly confirmed these earlier findings. Dustmann and Fabbri (2005) estimated a simple model using data from the 1979-2004 Labour Force Survey (LFS) and expanded the analysis to include females. They found that the wage gap for non-white immigrants was as large as -40%, though this varied with immigrants' region of origin. Dickens and McKnight (2008) estimated an unrefined model using data from the 1978-2003 Lifetime Labour Market Database (LLMDB) and found surprisingly large and negative wage gaps for all immigrants. In particular, they found a large wage penalty for white (European) immigrants, which is not in line with the UK or international literature.

considerably slows down after 30 years.

We thoroughly discuss the above issues in the remainder of this paper. In Section 2 we depict our data. In Section 3 and 4 we specify our empirical models and carefully discuss several identification issues. In Section 5 we summarize the results. In Section 6 we discuss our results before we conclude in Section 7.

2. Data

We use data from the Lifetime Labour Market Database (LLMDB). The LLMDB is derived from several administrative datasets linked together by a unique individual identifier, the national insurance number (NINo), which is similar to the US social security number. Individuals must apply for a NINo in order to pay tax (income tax, national insurance contributions, self-employment contributions, etc.), receive retirement pension, or claim social security benefits (e.g. unemployment benefit, incapacity benefit, sickness benefit, maternity benefit, child benefit, housing benefit, etc.). Whereas natives are automatically given a NINo just before they turn 16 years of age, which is derived from their child benefit number, immigrants typically apply for a NINo when they start interacting with the system, either by paying taxes or by claiming benefits. ("Natives" and "Immigrants" here and throughout the paper are respectively UK born and overseas born nationals.) Because individuals need to produce their NINo in every interaction with the system, the LLMDB effectively tracks individuals throughout their lifetime – and crucially, throughout their working lives.

The main advantage of the LLMDB, therefore, is that it is a rich, long and large longitudinal dataset. It has high levels of accuracy and relatively low levels of attrition (individuals only drop out of the sample if they neither pay taxes nor claim any benefit nor receive retirement pension for more than 12 months; they re-enter the sample when they again interact with the system). Our LLMDB sample comprises 647,068 individuals (a 1% random sample of NINo records) followed between the tax-years 1978 and 2006 (which run from April to March) resulting in 11,061,433 observations (a fresh cohort of individuals enters the data every year and is followed from then on). We restricted our sample to those aged 25 to 64, as is common in the immigration assimilation literature (though the results were robust when including those aged 16 to 64). We also restricted our sample to those earning between £100 and £1000000 in any one tax-year (this excludes the self-employed, for whom we do not observe earnings). We also restricted our sample to immigrants arriving from 1945 onwards, because the number of immigrants arriving previously was relatively very low and because

restricting the sample facilitates cohort modelling, which we discuss in Sections 3 and 4. Finally, we restricted our sample to those observed at least twice (though our results were robust to including individuals observed only once). Our final working sample therefore consists of 433,069 individuals, 45,309 of whom are immigrants, as shown in Table 1. The total number of observations is 5,373,086 and the average number of observations for natives (immigrants) is 15.76 (18.71). Figure 1 shows the distribution of natives and immigrants across tax-years.

The LLMDB contains well over 400 variables, including date of birth, date of death, age, gender, address, nationality, country which immigrants arrived from, immigrants' entry date,[3] immigrants' age at entry, number of jobs in the year, annual earnings per job, type of employment (employee or self-employed), number of weeks employed (unemployed) in the year, spells of unemployment, spells of receipt of benefits, benefit type, pension contributions, pension entitlements, etc. As is common with administrative data, the LLMDB does not contain information on education. In this paper we circumvent this issue to some extent by restricting our sample to those in work aged 25 to 64, who, we assume, have completed their education.[4] In addition, the LLMDB does not contain information on the immigrants' entry route (work permit, student visa, family reunification, etc.) or on their departure date. Since in this paper we focus on those already in work, their entry route is not crucial, though some limited information on such a route can be gauged from their first few interactions with the system. Similarly, since we are focusing on those in work, differentiating whether an immigrant left the labour force or left the country is not crucial here, although it might be very relevant elsewhere.

Table 1 shows that natives are older than immigrants, are more evenly spread across the country, are more likely to be employed and slightly less likely to be unemployed, and earn more on average. Immigrants at the very bottom of the earnings distribution earn less whereas those at the very top earn more than natives. Table 1 also shows that immigrants arrive young and many remain for several years. Immigrants predominantly come from the European

---

[3] The entry date is only recorded when the immigrant applies for a NINo, which depends on her individual circumstances and might not happen immediately upon arrival. Existing internal checks in the data and our own analysis suggest that the associated measurement error is fairly limited, especially after 1997.

[4] Although earnings models commonly control for education, there is an unresolved debate in the immigration literature about the interpretation of other coefficients in the model when controlling for education (Borjas 1999). Excluding education implies that we are comparing the earnings of immigrants and natives, and not the earnings of immigrants and natives with the same education level. This is important because the extent and quality of education varies across countries. Therefore, immigrants and natives with the same education may have different skills and compete for different jobs. Furthermore, immigrants across the education spectrum often suffer skill downgrading due to language or other labour market barriers.

Union (EU), Asia and the Middle East, and Africa, and there are marginally proportionately more white than non-white immigrants. Figure 2 confirms that these overall patterns persist across tax-years, though the proportion of non-whites increases over time. Interestingly, Table 1 shows that immigration intensified after the mid 1970s.[5] Figure 3 plots the inflow of immigrants by year of arrival and continent of nationality.[6] Most cohorts display a mix of white and non-white immigrants. For example, during the 1950s and 1960s, white EU immigrants (mainly Irish) and non-white immigrants from former colonies (India, Pakistan, Bangladeshi, South Africa, Nigeria, etc.) came to the UK to help with the post-war reconstruction effort. There were also other minority groups, such as Jews fleeing to the UK and refugees from Kenya, Uganda, etc. During the 1970s this overall trend continued, with increased EU immigration, especially after 1973, when the UK joined the EU. In this period there was also an increased inflow of immigrants from India (mainly because Gujarati Indians were expelled from Uganda) and from Vietnam and South East Asia. During the 1980s and 1990s white immigration increased proportionately more, following the enlargement of the EU when Greece (in 1981) and Portugal and Spain (in 1986) joined in, and following increases in the inflow of immigrants from North America, Australasia and Oceania. In that period non-white immigration, mainly from Africa, Asia and the Middle East, also increased. In the late 1990s and 2000s white immigration from the EU, North America, Australasia and Oceania continued to increase strongly, more dramatically after 2004, when ten Eastern European countries (A10) joined a further enlarged EU.[7] In this period non-white immigration from Africa, Asia and the Middle East increased sharply, and immigration from Central and South America also became more pronounced.

The above figures illustrate that disaggregation by continent of nationality and by cohort of arrival is another main advantage of the LLMDB. The LLMDB also permits disaggregation by small geographical levels.[8] This is in contrast with the more widely used Labour Force

---

[5] Although our sample period is between 1978 and 2006, the last cohort is four years short, by construction, since all those arriving in 2006 were dropped from the sample because they were only observed once.

[6] Although the LLMDB is more reliable after 1975, and more reliable still after 1997, Figure 3 plots the inflow from 1945 onwards. The overall number of immigrants is fairly reliable, but their continent of nationality was not always imputed. The rather large spikes in the series of "unknown" continent of nationality in Figure 3 reflects the introduction of new computing systems when inputting fields such as nationality of immigrants was not a priority. For example, the spike in 1975 reflects immigrants that arrived in the immediately preceding years (that is also true for the spikes in 1948 and 1997). Fortunately, this affects only three of our 13 cohorts of interest. Thus, despite this caveat in the data, it is possible to identify some overall patterns of immigration.

[7] Although the EU was successively enlarged at various points during our sample period, for consistency our definition of countries belonging to the EU throughout the whole period is that of 2006. We separately define the A10 countries, which are: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovenia and Slovakia, Malta and Cyprus.

[8] The ONS-defined geographical areas we use are: 409 Local Authority Districts, 49 counties and 12

Survey (LFS) – which is a rotating panel survey that interviews around 60,000 households with about 140,000 respondents every quarter available since 1992 (more comprehensive wages and hours worked data is available after 1997) – where immigration analysis across years and continents or below the regional level is not feasible due to sample size limitations.[9] Table 1 shows LFS variables. Dustmann and Fabbri (2005), using LFS data for a roughly comparable sample period, report descriptive statistics that are in line with our own descriptives here.

Table 1 shows that both the LLMDB and the LFS exhibit broadly similar patterns (we use the 1997-2007 sample period for this comparison). Women are slightly underrepresented in the LLMDB, perhaps reflecting their labour market participation decisions. The age distribution in both datasets is remarkably similar for natives, though a larger proportion of immigrants is younger in the LLMDB (note that we tabulate observations, not individuals). This is because a larger share of observations in the LLMDB is for immigrants (8.2% versus 7.7%) – and because the LLMDB better captures low paid immigrant workers, who tend to be younger. For example, the LLMDB might capture working foreign students and working illegal immigrants who might not have been captured in the LFS.[10] As a result, average earnings are lower in the LLMDB.

Importantly, the LLMDB records annual earnings (within the tax-year) – i.e. total annual earnings including any part-time and/or unemployment spells – whereas the LFS records weekly earnings in a given week, which are extrapolated for the year ignoring any part-time and/or unemployment spells (which are unknown). As a result, the LFS figures in Table 1 overestimate earnings, which are higher for every percentile of the distribution. The difference is larger at the bottom and smaller at the top of the distribution, confirming that the LLMDB captures more low paid workers (who either earn lower wages or work fewer hours). In particular, the LFS figures overestimate earnings for immigrants, who are more likely to be low paid, and thus the gap between natives and immigrants is less (more) persistent in the LFS (LLMDB), with immigrants earning more than natives up to the 20th (60th) percentile of the distribution. However, although earnings are consistently lower in the LLMDB, the average earnings trend over time is similar. (Detailed comparisons across years, available on request,

---

Government Regions (ONS 2003) (see Table 1).

[9] The recently available longitudinal LFS has an even smaller sample size and, furthermore, does not record county of residence or country of nationality prior to 2011.

[10] A small number of workers who earn too little or work too few hours to incur a national insurance contribution liability, for example those working part-time for very small employers, are not included in the LLMDB. Medium and larger employers are captured and their non-liable employees are included in the LLMDB. The LFS includes earnings for the self-employed, which are not recorded in the LLMDB (see above).

depict a very similar pattern of average earnings, percentile earnings, employment and unemployment rates across both datasets.) Dickens and McKnight (2008) carried out a similar analysis comparing the LLMDB and the ASHE and also concluded that annual earnings are lower in the LLMDB but that the trend of average earnings, and of selected percentiles of the earnings distribution, across both datasets is similar over time.

Finally, Table 1 shows that the regional distribution both in the LLMDB and in the LFS is remarkably similar for natives – and for immigrants, if those with unknown or abroad locations are excluded from the analysis. The distribution of immigrants' country of origin is also very similar in both datasets – again if those from unknown origin are excluded from the analysis. The distribution of immigrants' cohort of arrival is also very similar in both datasets (for a discussion of the spike at the 1975-1979 cohort in the LLMDB, see Footnote 6).

3. Cross-Section versus Longitudinal Analysis

We exploit the large sample size of the LLMDB to estimate a simple descriptive model of the immigrant-native earnings gap for each tax-year between 1978 and 2006, treating the data as 29 independent and separate cross-sections (Chiswick 2005; Lubotsky 2007):

$$E_i = a + bI_i + e_i \tag{1}$$

where E_i is log real earnings of individual i; I_i is an indicator variable that equals 1 if the individual is an immigrant; and e_i is the error term. The interpretation of our coefficient of interest is that immigrants on average earn $b\%$ more than natives – that is the immigrant-earnings gap is $b\%$.[11] We use generalized least square estimation and report robust White-corrected errors.

In each tax-year, we estimate this model for immigrants that are entering that tax-year; we then re-estimate the model controlling for the number of employed weeks in the tax-year. We also estimate the model for immigrants after 1 year of arrival, as well as after 10, 20 and 30 years of arrival, as shown in Figure 4 (and its counterpart Table A1).

At entry, Figure 4 suggests that immigrants earn roughly between 10% and 70% less than natives. After 10 years, this gap narrows to around zero for those arriving after the 1980s, with more recent cohorts doing better. This suggests that immigrants' earnings catch up with

---

[11] More precisely, immigrants on average earn $b^* = 100[\exp(b) - 1]$ more than natives. The closer b is to zero, the better it approximates b^*. Although some of our estimates are closer to zero than others, for simplicity we report b (and respectively β in Section 4) throughout the paper. Strictly speaking, immigrants on average earn b more than natives in logarithmic units (Halvorsen and Palmquist 1980).

natives' earnings after 10 years. After 20 years, Figure 4 suggests that immigrants earn roughly between 0% and 30% more than natives. After 30 years, the gap is positive and larger, though variation year-on-year is large, reflecting the smaller sample sizes and the different mix of immigrants arriving in each cohort (see Section 2). For example, the saw-toothed pattern in the mid and late 1990s might reflect the larger numbers of refugees from India, Kenya and Uganda in the mid and late 1960s. The sudden drop just before 2005 is, however, related to the new 1975 computing system (see Footnote 6); the same is true for the drop just before 1995 in the "After 20 years" series and the drop after 1997 in the "At Entry" series (see Section 2). Therefore the analysis for the 1965-1975 and 1997-1998 cohorts is best avoided as it can be misleading.

Another way to look at these results is to follow a particular immigrant cohort over time. For example, the immigrant-native earnings gap for immigrants that arrived in 1978 is -45.9% at entry, 4.4% after 10 years and 14% after 20 years of arrival (see Table A1); whereas the gap for those that arrived in 1977 is 12% after 10 years, 15.6% after 20 years and 21.7% after 30 years. Most coefficients are statistically significantly different from zero, although for earlier cohorts the coefficients are sometimes not significant.

However, these estimates might be biased. For example, if changes over time in unmeasured dimensions of immigrants' skills occurred during our sample period or if return migration of immigrants that are less able (along with other types of data attrition) occurred, then our estimates might be biased. Concerns with bias implied by various types of selectivity in the data, such as cohort bias and survivor bias, have occupied much of the literature for the last 30 years (Chiswick 1978, 1980; Carliner 1980; Borjas 1985, 1994 and 1999; Duleep and Regets 1997; Chiswick et al. 2005). This literature has long recognized that the ideal way to address such selectivity biases is to use sufficiently large and long longitudinal data – which has, however, been very scarce. Therefore, using the LLMDB to track immigrants over time when estimating the immigrant-natives earnings gap is a timely contribution, as it helps to prevent cohort bias and survivor bias. This is specially so because data attrition other than return migration in the LLMDB is fairly limited (see Section 2).

We thus re-estimate Equation (1) for a restricted sample of immigrants who stayed in the UK for 30 years or longer, as we show in Figure 5 (and its counterpart Table A2). The pattern of the immigrant-native earnings gap estimates is fairly similar to before, although the magnitude of the gap is consistently larger when using the sample of tracked immigrants. Since other sources of data attrition are not much of a concern in our data, the obvious

explanation here is return migration of the less able. The selectivity argument that more able, more motivated or more hardworking immigrants stayed on and were assimilated is in line with the larger more positive gap for the tracked sample. In other words, this suggests that the cross-section estimates in Figure 4 are downwards biased estimates of the immigrant-native earnings gap, and the estimates in Figure 5 give an indication of the extent of such survivor bias. For example, the gap after 20 years for immigrants that arrived in 1977 is now 26.1%, instead of 15.6% (see Tables A1 and A2).

When following particular immigrant cohorts over time, there is still great year-on-year variation. This suggests the presence of cohort bias to some extent in our estimates, perhaps reflecting changes over time in unmeasured dimensions of immigrants' skills, probably associated, at least in part, with the different mix of immigrants arriving in each cohort. However, in our cross-section analysis here, we do not control for other individual characteristics (other than the number of weeks employed in the year) that could proxy immigrants' skills; neither do we control for area or tax-year fixed effects. Thus, we now turn our attention to investigating whether such factors have an effect on our immigrant-native earnings gap estimates.

4. Economic Assimilation

Our estimates in Section 3 suggest the existence of an immigrant-native earnings gap in the UK between 1978 and 2006, which is quite sizeable for some groups of immigrants. However, such unconditional earnings gap estimates need to be proved robust when accounting for the effect of other individual characteristics (such as gender, age, continent of nationality, years since immigration, etc.) on earnings. We now account for this by estimating a conditional immigrant-native earnings gap model (conditional on such individual characteristics). Those unconditional estimates also need to be proved robust to when the specific effect of areas and tax-years on earnings is accounted for. We now account for this by exploiting the longitudinal dimension of our data. This way we can disentangle cohort and assimilation effects.

We use what is now a standard immigrant economic (earnings) assimilation model in the literature (Chiswick 1980; Bell 1997; Borjas 1999; Lubotsky 2007):

$$E_{iat} = \alpha + \beta I_i + \gamma Z_{iat} + \lambda X_{iat} + \lambda^I I_i X_{iat} + f_a + f_t + \varepsilon_{iat} \tag{2}$$

where E_{iat} is log real earnings of individual $i = 1,...,433069$ in area $a = 1,...,49$ and time $t = 1978,...,2006$; I_i is as before an indicator variable that equals 1 if the individual is an

9

immigrant; Z_{iat} is the number of "years since immigration" for immigrants and is zero for natives; X_{iat} is a vector of observable individual characteristics; f_a is area fixed effects; f_t is time fixed effects; and ε_{iat} is the error term.

The interpretation of our first coefficient of interest is that immigrants on average earn $\beta\%$ more than natives at entry – that is the immigrant-native earnings gap at entry is $\beta\%$ – when compared with natives with similar individual characteristics (see Footnote 11). The interpretation of our second coefficient of interest is that each year spent in the UK on average increases the earnings of immigrants by $\gamma\%$. We can also use the model's estimates to calculate the earnings gap after arrival, as well as the immigrant earnings growth rate and the rate of earnings convergence between immigrants and natives.[12]

We model area fixed effects using county dummies (see Section 2). This way we remove any permanent differences across counties and make them equally attractive. In other words, we control for specific factors in a county (such as more schools, more housing, lower prices, etc.) that may make it more attractive to immigrants or natives or both. This enables us to separately account for the effect of county specific time invariant factors on earnings. We model time fixed effects using tax-year dummies (see Section 2). This way we control for the effect of tax-year specific macroeconomic effects (such as seasonal shocks, national and international macroeconomic shocks, etc.). This enables us to separately account for the effect of tax-year specific factors on earnings. Controlling for area and time fixed effects in this flexible manner (across counties and tax-years) is an improvement to the existing UK earnings gap literature.[13]

We also control for observable individual characteristics such as sex, age, age squared, number of employed weeks in the year and number of jobs in the year (see Table 1). We allow

---

[12] The immigrant earnings growth rate is γ times the number of years since arrival; whereas the immigrant-native earnings convergence rate is the immigrant earnings growth rate added to the difference between the coefficient of age, times the number of years after arrival, for immigrant and natives (Borjas 1999). The earnings gap after arrival is the immigrant earnings growth rate added to β and λ^I (the coefficients of the number of employed weeks in the year and the number of jobs in the year are timed by their respective averages in the immigrant sample from Table 1).

[13] The available models in this literature do not control for area fixed effects, except for Dustman and Fabbri (2005), who include region (not county) fixed effects. Here we model area fixed effects using 49 counties instead of 12 regions, which is a more flexible approach. On a related point, the assumption that area and time fixed effects are the same for immigrants and natives is restrictive, since macroeconomic conditions most likely affected both groups differently. This is particularly so if immigrants are concentrated in the low or high end of the earnings distribution, as suggested by Table 1 (see Section 2); or if the earnings distribution changed over time (LaLonde and Topel 1992; Butcher and DiNardo 2002). However, this is a common restriction in the literature (Borjas 1999).

the effect of these variables to differ between natives and immigrants. This enables us to separately account for the effect of such characteristics on the earnings of each group. For example, this way we account for earnings differentials due to workers being younger in addition to being immigrants. Although we do not observe experience, we control for age, which, albeit imperfectly, captures overall experience to a certain extent (Borjas 1999). In addition, we control for "years since immigration" and its squared value to account for the effect of extra time spent in the UK on earnings. Since "years since immigration" is often identical to experience in the UK, as most immigrants arrived as adult workers, this enables us to gauge the weight that employers attach to experience gained in the UK.[14]

We use generalized least square estimation and correct for intragroup serial correlation, as standard errors are assumed to be independent across groups of individuals but not within groups (i.e. standard errors are not assumed to be independent for a particular individual over time).

5. Results

The first column in Table 2 (and of its counterpart Table A3) shows a significant -0.462 estimate of β. This estimate suggests that immigrants on average earn 46.2% less than comparable natives at entry. This is fairly close to our cross-section estimates for cohorts entering in the 1980s and 1990s in Section 3, though larger than for those cohorts entering in the early 2000s. This suggests that the earnings gap at entry is relatively robust to controlling for other individual characteristics and area and time fixed effects.

This immigrant-native earnings gap at entry narrows to -0.281 after 10 years, -0.042 after 20 years and 0.104 after 30 years. This suggests that immigrants' earnings catch up with natives' earnings a little after 20 years, which is more than 10 years later than our cross-section estimates suggested in Section 3. This indicates that other individual characteristics and area and time fixed effects play a role in determining the immigrant-native earnings gap after entry.

---

[14] We experimented with controlling for "age at entry" – which captures the human capital endowment at arrival and is particularly important for identifying immigrants who arrived as children and thus have labour market characteristics of natives – instead of age and age squared (due to collinearity with "years since immigration"), but this did not alter our main results (also see Borjas 1994 and 1999). In our data sample, most immigrants arrived as adults; around 10% entered as children, and the vast majority of these arrived before 1969 from ex-colonies and from Europe. Our results were also robust to including extra controls such as receipt of maternity benefit, child benefit, incapacity benefit, sickness benefit, etc. Most of these, however, are not consistently available for the entire sample period, since they were introduced or changed at various points in time.

5.1 Cohort Effects

Our immigrant-native earnings gap estimates in Section 5 pool together very diverse groups of immigrants who differ widely in several individual characteristics. One such important characteristic is their cohort of arrival. If all cohorts displayed a similar mix of immigrants in terms of continent of nationality, English proficiency, skills, motivation, work ethics, etc., and if the economic conditions in the UK were the same throughout the whole sample period (along with natives' attitudes to migration), then there would be no reason to expect earnings gap estimates to vary with cohort of arrival. We largely account for local and national macroeconomic conditions in the UK affecting earnings when we control for county and tax-year fixed effects. However, it is still possible that the earnings gap estimates are affected by immigrants' individual characteristics specific to their cohort of arrival. This is an identification issue that has played an important role in the immigration literature (Borjas 1985 and 1999). This literature has long recognized that a sufficiently large and long longitudinal data – a very scarce kind of dataset – is the ideal way to estimate such cohort effects separately from assimilation effects. In addition, producing estimates by cohort of arrival is, of course, informative in itself.

Thus, we now exploit the large and long nature of our longitudinal data to re-estimate Equation (2) including an explicit indicator for 13 five-year cohorts of arrival as a proxy for such "cohort" individual characteristics. This way we account for characteristics that vary across cohorts such as unmeasured dimensions of immigrants' skills or return migration of immigrants that are more or less able. Estimating the earnings gap in such a flexible manner across cohorts of arrival is an important contribution to the existing UK earnings gap literature, where such estimates are largely unavailable. This is in contrast with the international literature, where considerable effort has been put into estimating such cohort effects (Borjas 1985 and 1999).[15]

The second column in Table 2 (and of its counterpart Table A3) shows significant cohort effect estimates. These estimates suggest that the immigrant-native earnings gap is more negative for earlier cohorts of arrival. For example, immigrants that arrived in the post-war period on average earned 55% to 60% less than comparable natives at entry; whereas immigrants that arrived respectively during the 1970s and 1980s and then during the mid 1990s to the mid 2000s earned around 40% and 30% less than comparable natives at entry.

---

[15] The available models in the UK literature do not control for cohort of arrival, except for Bell (1997), who includes an indicator for 4 ten-year-cohorts. Here we include an indicator for 13 five-year-cohorts for a sample period that is 15 years longer, which is a more flexible approach.

This contrasts with estimates between -15% to -35% for the US on a roughly comparable model, which however uses the wage rate, instead of annual earnings, different controls and data from the 1970, 1980 and 1990 Census (Borjas 1995 Table 5 and Borjas 1999 Table 4). Bell (1997) also estimated a comparable model for the UK using household survey data for the 1970s and 1980s, and found successively smaller cohort estimates for some groups of immigrants but not for others (also see Footnote 15).

The notable exception is those who arrived in the mid to late 1960s, whose earnings gap was smaller than those arriving two decades before or after. As discussed in Sections 2 and 3, this period witnessed relatively larger numbers of refugees from India, Kenya and Uganda, and the effect of this inflow on our estimates is not obvious. For example, if the individuals selected into our sample are the most able and motivated of all such refugees who immediately found work, our estimates might be less negative (see Section 3). Or if most such refugees are selected out of our sample because they did not immediately find work or were not allowed to work, our estimates might be less negative if such nationalities are typically lower earners. Since we do not observe entry route, we cannot determine whether the relative proportion of refugees increased or decreased from one cohort to another in our sample.[16]

Table 2 shows that the immigrant-native earnings gap for those arriving in 1945-1949 narrows from -0.589 at entry to -0.418 after 10 years, to -0.175 after 20 years and to -0.015 after 30 years. In contrast, the immigrant-native earnings gap for those arriving in 2000-2004 narrows from -0.300 at entry to -0.129 after 10 years, to 0.114 after 20 years and to 0.274 after 30 years. This suggests that immigrants arriving later not only have a smaller earnings gap at entry, but also have a faster earnings catch-up over time. While the convergence took over 30 years for those entering in 1945-1949, it only took half as long for those entering in 2000-2004. Incidentally, the convergence for the pooled sample was a little over 20 years, which is roughly the average between the earlier and later cohorts (see Section 5 and columns 1 and 2 of Table 2). Figure 6 illustrates the earnings convergence over time for selected cohorts.

5.2 Nationality Effects

Our immigrant-native earnings gap estimates in Section 5.1 still pool together very diverse

---

[16] For the single-year-cohort 2005, the gap estimate is more negative than that for the two previous cohorts. This could be due to relatively more low earners from A10 countries entering the sample, or it could be due to this cohort being four years shorter (this cohort constitutes a smaller sample than the other five-year-cohorts, for which furthermore, there are at most two usable observations for each individual, as discussed in Section 2). This suggests that estimates for this cohort should be treated with caution.

groups of immigrants who differ widely in several individual characteristics (such as English proficiency, work ethics, skills (formal education) transferability, etc.) and might not always be perfect labour substitutes. This is because all cohorts display a mix of nationalities, and thus very different immigrants are pooled together in each cohort. We thus re-estimated Equation (2) including an explicit indicator for 8 continents of nationality as a proxy for such "group" individual characteristics. Producing estimates by continent of nationality is, of course, informative in itself, as recognized in the existing UK and international literature (Chiswick 1980; Borjas 1994; Butcher and DiNardo 2002; Dustmann and Fabbri 2005; Dickens and McKnight 2008). In addition, estimates by continent of nationality can be loosely compared with estimates by cohort of arrival, since in most cohorts there are one or two dominant continents of nationality. Thus, a comparison between estimates by cohort of arrival and estimates by continent of nationality provides a further insight into the immigrant-native earnings gap phenomenon from a different angle.

The third column in Table 2 (and of its counterpart Table A3) shows significant continent of nationality estimates. These estimates suggest that the immigrant-native earnings gap is more negative for non-white than for white immigrants. For example, non-white immigrants from Africa, Asia and the Middle East, and Central and South America earned around 50% less than comparable natives at entry; whereas white immigrants from Europe and the EU earned around 40% less, immigrants from Australasia and Oceania earned around 30% less, and immigrants from North America earned only 16% less than comparable natives at entry. Interestingly, the variation in the entry gap estimates across non-whites (51% to 52%) is much smaller than the variation in the entry gap estimates across whites (16% to 41%). The notable exception is white immigrants from the A10 countries, who earned around 50% less than comparable natives at entry.[17]

Table 2 shows that the immigrant-native earnings gap for Africans narrows from -0.511 at entry to -0.365 after 10 years, to -0.147 after 20 years and to -0.016 after 30 years. In contrast, the immigrant-native earnings gap for North Americans narrows from -0.160 at entry to -0.014 after 10 years, to 0.204 after 20 years and to 0.335 after 30 years. This suggests that North Americans not only have a smaller earnings gap at entry, but also a faster earnings convergence. While Africans took over 30 years to catch up with natives' earnings, North

---

[17] As before, this could be due to relatively more low earners from the large inflow of A10 workers entering the sample in 2004 and 2005. As in the case of estimates by cohort of arrival, for many A10 workers there are at most two usable observations for each individual, which suggests that estimates for this nationality should be treated with caution.

Americans took just over 10 years. Incidentally, the convergence for the pooled sample was a little over 20 years, in line with the presence of proportionately more lower earners in the sample, as indicated in Figures 2 and 3 (see Section 5). Figure 7 illustrates the earnings convergence over time for all continents.

A comparison between these continent of nationality estimates and our earlier cohort of arrival estimates from Section 5.1 suggests that, broadly speaking, the immigrant-native earnings gap at entry, and the subsequent catch-up, is less negative in cohorts with proportionately more white immigrants. This pattern is broadly true when comparing earlier post-war cohorts, which witnessed non-white immigration from former colonies, and more recent cohorts, which witnessed several waves of white post-EU enlargement immigration. This pattern is also true when comparing more recent cohorts, though the trend is less obvious and more subtle. For example, this pattern can be seen when comparing the immediately adjacent cohorts of 1985-1994 and 1995-2004. In the 1985-1994 cohorts, when non-white immigration increased, the earnings gap at entry is -0.440 to -0.406 (though this period also coincides with greater white immigration following the enlargement of the EU in the mid 1980s). In contrast, in the 1995-2004 cohorts, when white immigration increased, the earnings gap at entry is -0.305 to -0.300 (though this period also witnesses greater immigration from Asia and the Middle East as well as from Africa). Thus, it can be argued that the entry earnings gap fell over time (except for the 1965-1974, as discussed in Section 5.1) but that this fall slowed and reversed in 1985-1994, when non-white immigration increased. Likewise, it can be argued that the entry earnings gap fell sharply in 1995-2004, when white immigration increased. Bell (1997) estimated a comparable model using 1973-1992 UK GHS survey data, and also found that the gap at entry, and the subsequent catch-up, is more negative for non-white immigrants (see Footnotes 2 and 15). Dustmann and Fabbri (2005) estimated a simpler model using data from the 1979-2004 UK LFS survey data and also found that the gap for non-white immigrants was more negative.

5.3   Earnings Growth and Rate of Earnings Convergence

Table 2 also shows the immigrant earnings growth rate and the immigrant-native earnings convergence rate over time in the pooled sample, by cohort of arrival and by continent of nationality. The first column in Table 2 shows that immigrant earnings growth in the pooled sample is 0.126 after 10 years. This suggests that newly arrived immigrants earn on average 12.6% less than otherwise comparable immigrants that arrived 10 years earlier. Immigrant

earnings growth is 0.233 after 20 years, 0.322 after 30 years, and 0.392 after 40 years. Thus earnings growth is fastest in the first 10 years, and is much slower after 30 years, when it tends to a plateau.

The associated immigrant-native earnings convergence rate is 0.333 after 10 years. This suggests that the earnings of immigrants increase 33.3% faster than the earnings of natives in the first 10 years. The rate of earnings convergence is 0.573 after 20 years, 0.718 after 30 years, and 0.770 after 40 years. Thus the rate of convergence is also fastest in the first 10 years, and is much slower after 30 years, when it also tends to a plateau.

The second column in Table 2 shows that immigrant earnings growth in the model where we control for cohort of arrival is larger than in the pooled model over time (0.176 versus 0.126), though growth is still fastest in the first 10 years and much slower after 30 years. In contrast, the third column in Table 2 shows that immigrant earnings growth in the model where we control for continent of nationality is smaller than in the pooled model over time (0.112 versus 0.126), although, again, growth is still fastest in the first 10 years and much slower after 30 years. Therefore, the estimates of immigrant earnings growth rate are fairly robust across all three models, which is reassuring.

The same is true for estimates of the rate of earnings convergence. They are fairly robust across all three models, although they are now smaller both in the cohort of arrival model (0.326 versus 0.333) and in the continent of nationality model (0.305 versus 0.333) (interestingly, they are larger after 30 and 40 years in the cohort of arrival model only). Like before, the earnings convergence rate is fastest in the first 10 years, and it is much slower after 30 years.

As no such estimates for the UK are available to compare our estimates with, we attempt a comparison with US estimates. Our estimates contrast with an immigrant earnings growth rate estimate of 0.076 after 10 years and an earnings convergence rate estimate of 0.060 after 10 years for the US on a roughly comparable model (Borjas 1999 Table 4). Thus, both figures are smaller for the US, although the entry earnings gap was closer to our estimates, as discussed above. There are many reasons why our estimates are larger, including differences in the labour market and in the immigrant population composition, as well as differences in model specification, data type and sample period. The US model uses the wage rate, instead of annual earnings, different controls and data from the 1970, 1980 and 1990 Census. Other obvious differences include the fact that the US has had substantial non-white low skilled immigration whereas the UK has had comparatively larger white highly skilled immigration

(Borjas 1994; Dustman and Glitz 2005). In addition, perhaps because the US labour market is more flexible, low paid immigrants in the UK have a more negative earnings gap at entry (see Section 5.1) and thus have faster earnings growth (Chiswick et al. 2008). Further, such low paid workers in the UK might be more skilled and hence might overcome the usual earnings and occupation downgrading they suffer at entry more quickly (Friedberg 2001; Manacorda et al. 2006).

5.4   Robustness Checks

Our models above control for weeks worked, which ensures that we are comparing immigrants and natives that work the same number of weeks in the year. However, even when immigrants and natives work the same number of weeks in the year, they could still work different hours within the week. For example, if natives, say, earn higher earnings, but also work longer hours, this could explain part of an earnings gap in favour of natives. In other words, uncontrolled different hours worked per week could potentially bias the estimate of β in our models above.

As we are unable to control for hours worked in our models, since this variable is unavailable in the LLMDB, we use data from the LFS to gauge the direction of such potential bias (see Section 2). Table 1 shows that immigrants work around 20 minutes longer than natives in the week of reference on average. A simple average test shows that this difference is statistically significant (t = -9.44) and hence suggests that our β estimates above are biased. This suggests that some of the immigrant-native earnings gap in Table 2 is explained by the difference in hours worked.

One way to gauge the magnitude of this potential bias is to contrast estimates for men and women. The bias is potentially larger for women, for whom the difference in hours worked is larger (immigrant women work around 100 minutes longer, with the associated t-test = -31.92). Indeed, columns 4 and 5 in bold of Table A3 show that the immigrant-native earnings gap is significant and larger for women (-0.573) than for men (-0.449), with the latter being fairly close to the pooled estimate (-0.462). We thus argue that the magnitude of any associated bias here is small.

The place where such potential bias would be smallest is for men working full time. This is because the difference in hours worked is smallest among men who work a full year. Immigrant men who worked a full year (12 months continuously from March to April in the LFS, to coincide with the LLMDB tax-year) work around 15 minutes longer (with the

associated t-test = -2.14). This is a small difference, and one which is not significant at the 10% level. We re-estimate our model separately for men and women working 52 weeks per year. The idea here is that both immigrants and natives working a full year are perhaps also more likely to work a full week. Producing estimates for "full timers" is, of course, informative in itself. Once again the immigrant-native earnings gap is significant and larger for women (-0.493) than for men (-0.441). The difference in magnitude between the latter and the earlier one (-0.449) is very small, which suggests, in turn, that the magnitude of any associated bias is small.

6. Discussion

We estimated the immigrant-native earnings gap at entry and over time for the UK between 1978 and 2006. We initially utilized our long and large longitudinal data as a succession of independent cross-sections to estimate a simple descriptive immigrant-native earnings gap model. However, because of concerns with bias implied by various types of selectivity in the data, such as cohort bias and survivor bias, we re-estimated our simple model using a restricted sample of immigrants who stayed in the UK for 30 years or longer. The pattern of both our unrestricted and restricted estimates was fairly similar, although the magnitude of the gap was consistently larger when using the restricted sample of tracked immigrants. Since other sources of data attrition are not much of a concern in our data, the obvious explanation here was return migration of the less able. The selectivity argument that more able, more motivated or more hardworking immigrants stayed on and were assimilated is in line with the larger more positive gap for the tracked sample. In other words, the unrestricted estimates were downwards biased and a comparison with the restricted estimates indicated the extent of such survivor bias.

Nonetheless, the restricted estimates still displayed large year-on-year variation. This suggested the presence of cohort bias to some extent in our estimates, perhaps reflecting changes over time in unmeasured dimensions of immigrants' skills, probably associated, at least in part, with the different mix of immigrants arriving in each cohort. As our simple cross-section unconditional earnings gap cross-section model did not control for other individual characteristics that could proxy immigrants' skills, or for area or tax-year fixed effects, we then exploited the longitudinal nature of our data to estimate a more complete conditional earnings gap model to disentangle cohort and assimilation effects.

Our estimates showed that the earnings gap at entry was relatively robust to controlling for

18

other individual characteristics and area and time fixed effects. However, these extra controls did play a marked role in the estimates of the immigrant-native earnings gap after entry, which was then less favourable.

We drew two main conclusions from our longitudinal model estimates. Firstly, more recent cohorts fare better than earlier ones at entry. For example, immigrants that arrived in the post-war period on average earned 55% to 60% less than comparable natives at entry; whereas immigrants that arrived during the 1970s and 1980s respectively and then during the mid 1990s to the mid 2000s earned around 40% and 30% less than comparable natives at entry.

Secondly, the earnings of immigrants from more recent cohorts catch up faster with natives' earnings. For example, the immigrant-native earnings gap for those arriving in 1945-1949 narrows from -59% at entry to -42% after 10 years, and to -1.5% after 30 years. In contrast, the immigrant-native earnings gap for those arriving in 2000-2004 narrows from -30% at entry to -13% after 10 years, and to 27.5% after 30 years. While the convergence took over 30 years for those entering in 1945-1949, it only took half of as long for those entering in 2000-2004.

The associated immigrant earnings growth rate estimates suggested that newly arrived immigrants earn around 12.5% less than otherwise comparable immigrants that arrived 10 years earlier and 32% less than those that arrived 30 years earlier. Thus, earnings growth is fastest in the first 10 years, and it considerably slows down after 30 years when it tends to a plateau. The associated immigrant-native earnings convergence rate estimates suggested that the earnings of immigrants increase 33% faster than the earnings of natives in the first 10 years after arrival, and 72% faster in the first 30 years, when earnings convergence also slows down and tends to a plateau.

7. Conclusion

We exploit a large and long longitudinal dataset to estimate the immigrant-native earnings gap at entry and over time for the UK between 1978 and 2006. That is, we attempt to separately estimate cohort and assimilation effects for the UK for these three decades. We also estimate the associated immigrant earnings growth rate and immigrant-native earnings convergence rate over time.

Our estimates suggest that immigrants from more recent cohorts fare better than earlier ones at entry. For example, whereas immigrants that arrived in the post-war period earned nearly 60% less than comparable natives at entry, those arriving in the early 2000s earned

only 30% less at entry. Furthermore, our estimates suggest that the earnings of immigrants from more recent cohorts catch up faster with natives' earnings. While the convergence took over 30 years for those entering in the post-war, it only took half as long for those entering in the early 2000s.

Our estimates also suggest that newly arrived immigrants earn around 12.5% less than otherwise comparable immigrants that arrived 10 years earlier and 32% less than those that arrived 30 years earlier. Thus, earnings growth is fastest in the first 10 years, and it considerably slows down after 30 years. Finally, our estimates suggest that the earnings of immigrants increase 33% faster than the earnings of natives in the first 10 years after arrival, and 72% faster in the first 30 years, when earnings convergence also slows down.

These new results are an important contribution to the UK literature, where assimilation and cohort effect estimates are largely unavailable. These new results are also an important contribution to the international literature, which has long recognized that a sufficiently large and long longitudinal data is the ideal way to estimate cohort effects separately from assimilation effects. By applying a thorough and careful approach in estimating these effects to one such a dataset, we provide a timely and important contribution to the literature.

8. References

Bell, B. D. (1997): "The Performance of Immigrants in the United Kingdom: Evidence from the GHS," *The Economic Journal*, 107, 333-44.

Borjas, G. J. (1985): "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463-489.

— (1994): "The Economics of Immigration," *Journal of Economic Literature*, 32, 1667-1717.

— (1995): "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?," *Journal of Labor Economics*, 13, 201-245.

— (1999): "The Economic Analysis of Immigration," in *Handbook of Labor Economics, Vol. 3a*, ed. by O. Ashenfelter, and D. Card. Amsterdam: North-Holland, 1697-1760.

Butcher, K. F., and J. DiNardo (2002): "The Immigration and Native-Born Wage Distributions: Evidence from United States Censuses," *Industrial and Labor Relations Review*, 56, 97-121.

Carliner, G. (1980): "Wages, Earnings and Hours of First, Second and Third Generation American Males," *Economic Inquiry*, 18, 87-102.

Chiswick, B. R. (1978): "The Effects of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86, 897-921.

— (1980): "The Earnings of White and Coloured Male Immigrants in Britain," *Economica*, 47, 81-87

Chiswick, B. R., A. T. LE, and P. W. Miller (2008): "How Immigrants Fare across the Earnings Distribution in Australia and the United States," *Industrial and Labor Relations Review*, 61, 353-373.

Chiswick, B. R., Y. L. Lee, and P. W. Miller (2005): "Immigrant Earnings: A Longitudinal Analysis," *Review of Income and Wealth*, 51, 485-503.

Cortes, K. E. (2004): "Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States," *The Review of Economics and Statistics*, 86, 465-480.

Dickens, R., and A. McKnight (2008): "Assimilation of Migrants into the British Labour Market," *CASE Discussion Paper*, 133.

Duleep, H., and M. Regets (1997): "Measuring Immigrant Wage Growth Using Matched Cps Files," *Demography*, 34, 239-249.

Dustmann, C., and F. Fabbri (2005): "Immigration in the British Labour Market," *Fiscal Studies*, 26, 423-470.

Dustmann, C., and A. Glitz (2005): *Immigration, Jobs and Wages: Theory, Evidence and Opinion*. CEPR and CReAM.

Friedberg, R. M. (2001): "The Impact of Mass Migration on the Israeli Labor Market," *The Quarterly Journal of Economics*, November, 1373-1408.

Halvorsen, R., and R. Palmquist (1980): "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 70, 474-475.

Lalonde, J. R., and H. R. Topel (1992): "The Assimilation of Immigrants in the U.S. Labor Market," in *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*, ed. by G. Borjas, and R. Freeman. Chicago: University of Chicago Press, 67-92.

Lubotsky, D. (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy*, 115, 820-867.

Manacorda, M., A. Manning, and J. Wadsworth (2006): "The Impact of Immigration on the Structure of Male Wages: Theory and Evidence from Britain," *IZA Discussion Paper*, 2352.

ONS (2003): "United Kingdom: Local Authority Districts, Counties and Unitary Authorities, 1998," *ONS Geography GIS & Mapping Unit*.

Table 1 - DESCRIPTIVE STATISTICS						
VARIABLES	**LLMDB**		**LLMDB**		**LFS**	
	April 1978 - March 2007		**April 1997 - March 2007**		**January 1997 - March 20**	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
I - POPULATION VARIABLES						
% aged:						
25 to 34 years old	31.34%	43.83%	29.47%	43.24%	29.06%	36.59%
35 to 64 years old	68.66%	56.17%	70.53%	56.76%	70.94%	63.41%
% of women	43.43%	44.85%	46.67%	44.37%	48.47%	47.51%
% from:						
EU (except A10)	-	32.20%	-	30.99%	-	25.66%
A10	-	4.03%	-	5.41%	-	5.43%
Europe (except EU)	-	3.15%	-	3.64%	-	2.65%
Asia and Middle East	-	20.89%	-	22.38%	-	27.94%
North America	-	6.09%	-	5.04%	-	4.63%
Latin America	-	3.24%	-	3.49%	-	6.59%
Africa	-	13.80%	-	16.16%	-	21.39%
Australasia and Oceania	-	7.30%	-	6.39%	-	4.74%
Unknown	-	9.31%	-	6.51%	-	na
Average age at arrival	-	23.03	-	24.43	-	na
Average nb of years since immigration	-	14.73	-	13.30	-	na
% with lenght of immigration						
0 to 1 years	-	8.64%	-	11.15%	-	na
2 to 3 years	-	10.17%	-	12.54%	-	na
4 to 5 years	-	8.48%	-	9.73%	-	na
6 to 10 years	-	16.96%	-	17.06%	-	na
11 to 15 years	-	15.24%	-	14.02%	-	na
16 to 20 years	-	11.70%	-	10.46%	-	na
over 20 years	-	28.82%	-	25.04%	-	na
% arrived during:						
1945-1949	-	3.87%	-	0.04%	-	1.03%
1950-1954	-	2.63%	-	0.46%	-	1.77%
1955-1959	-	4.54%	-	1.69%	-	3.45%
1960-1964	-	5.79%	-	2.84%	-	7.23%
1965-1969	-	6.05%	-	3.38%	-	8.57%
1970-1974	-	4.38%	-	2.79%	-	9.16%
1975-1979	-	17.45%	-	11.77%	-	7.61%
1980-1984	-	9.62%	-	8.48%	-	5.57%
1985-1989	-	12.72%	-	13.47%	-	7.69%
1990-1994	-	10.54%	-	14.22%	-	9.10%
1995-1999	-	10.59%	-	18.95%	-	13.49%
2000-2004	-	10.76%	-	19.92%	-	13.26%
2005	-	1.08%	-	1.99%	-	1.39%
% located in:						
East Midlands	7.64%	4.26%	7.68%	4.28%	7.66%	4.79%
East of England	9.50%	7.28%	9.55%	7.62%	9.69%	9.12%
London	8.73%	29.87%	9.03%	35.14%	9.61%	41.76%
North East	4.46%	2.08%	4.35%	1.88%	4.32%	1.31%
North West	11.72%	5.54%	11.56%	5.57%	11.97%	5.57%
Northern Ireland	2.39%	1.26%	2.51%	1.29%	2.51%	1.22%
Scotland	9.43%	5.16%	9.39%	4.78%	9.03%	3.67%
South East	13.73%	12.37%	13.93%	12.87%	13.99%	13.87%
South West	8.79%	5.25%	8.76%	5.09%	8.51%	5.24%
Wales	4.87%	2.17%	4.79%	1.95%	4.87%	1.80%
West Midlands	9.16%	5.10%	9.10%	5.26%	9.20%	6.90%
Yorkshire and the Humber	8.68%	4.12%	8.66%	4.13%	8.66%	4.74%
Unknown or Abroad	0.90%	15.55%	0.69%	10.15%	-	-
					(continues)	

Table 1 - DESCRIPTIVE STATISTICS (continued)						
VARIABLES	**LLMDB**		**LLMDB**		**LFS**	
	April 1978 - March 2007		**April 1997 - March 2007**		**January 1997 - March 20**	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
II - LABOUR MARKET VARIABLES						
% in work:						
1 to 25 weeks in the year	16.74%	26.05%	18.84%	30.33%	na	na
26 to 50 weeks in the year	13.93%	18.96%	15.45%	20.95%	na	na
51 to 52 weeks in the year	69.09%	54.76%	65.55%	48.54%	na	na
Average number of employed weeks in the year	43.18	38.52	42.10	36.40	na	na
Average number of unemployed weeks	1.37	1.39	1.03	1.17	na	na
Average number of jobs in the year	1.33	1.53	1.44	1.66	na	na
Average total actual hours worked in the week	na	na	na	na	36.75	37.10
5th percentile of the log real earnings distribution	7.60	7.27	7.65	7.25	9.22	9.15
10th percentile of the log real earnings distribution	8.26	7.97	8.35	7.99	9.39	9.34
20th percentile of the log real earnings distribution	8.87	8.65	8.93	8.68	9.59	9.57
30th percentile of the log real earnings distribution	9.26	9.09	9.30	9.12	9.73	9.74
40th percentile of the log real earnings distribution	9.50	9.40	9.55	9.43	9.86	9.88
50th percentile of the log real earnings distribution	9.68	9.64	9.74	9.68	9.99	10.03
60th percentile of the log real earnings distribution	9.84	9.86	9.91	9.90	10.12	10.17
70th percentile of the log real earnings distribution	9.99	10.06	10.08	10.13	10.26	10.32
80th percentile of the log real earnings distribution	10.16	10.29	10.27	10.37	10.42	10.51
90th percentile of the log real earnings distribution	10.41	10.62	10.52	10.74	10.65	10.80
Average of the log real earnings distribution	9.48	9.45	9.55	9.50	10.00	10.04
Standard deviation of the log real earnings distribution	0.95	1.13	0.99	1.18	0.53	0.61
Number of observations	5053659	319427	1935699	172466	507606	42230
Number of individuals	387760	45309	277532	35415	na	na
Average number of times an individual is observed	15.76	18.71	24.53	25.16	na	na
% observations per year:						
1978	2.89%	1.69%	-	-	-	-
1979	3.16%	2.02%	-	-	-	-
1980	3.07%	1.97%	-	-	-	-
1981	3.09%	1.96%	-	-	-	-
1982	3.08%	1.95%	-	-	-	-
1983	3.07%	1.95%	-	-	-	-
1984	3.14%	2.05%	-	-	-	-
1985	3.08%	2.03%	-	-	-	-
1986	3.11%	2.08%	-	-	-	-
1987	3.11%	2.14%	-	-	-	-
1988	3.26%	2.35%	-	-	-	-
1989	3.34%	2.52%	-	-	-	-
1990	3.42%	2.67%	-	-	-	-
1991	3.38%	2.74%	-	-	-	-
1992	3.42%	2.90%	-	-	-	-
1993	3.44%	3.02%	-	-	-	-
1994	3.50%	3.19%	-	-	-	-
1995	3.56%	3.32%	-	-	-	-
1996	3.56%	3.45%	-	-	-	-
1997	3.68%	3.67%	9.62%	6.80%	10.51%	9.23%
1998	3.73%	3.91%	9.75%	7.24%	11.55%	10.50%
1999	3.81%	4.30%	9.95%	7.96%	11.11%	10.15%
2000	3.86%	4.69%	10.08%	8.69%	10.56%	9.27%
2001	3.88%	5.00%	10.13%	9.26%	7.72%	7.30%
2002	3.88%	5.41%	10.13%	10.01%	10.14%	10.27%
2003	3.90%	5.91%	10.18%	10.94%	9.73%	9.86%
2004	3.83%	6.52%	9.99%	12.08%	9.23%	9.54%
2005	3.90%	7.43%	10.17%	13.75%	8.87%	10.23%
2006	3.83%	7.16%	10.00%	13.26%	8.44%	10.79%
Average number of observations per year	174264	11015	193570	17247	50761	4223

Source: Lifetime Labour Market Database and Labour Force Survey.
(1) Sample includes those aged 25 to 64 employed and earning between £100 and £1000000 in the year.
(2) LFS figures are courtesy of the DWP.

Model	estimate	estimate	estimate	
Immigrant-Native Earnings Gap at Entry				
Pooled	-0.462			
By Cohort of Arrival				
1945-1949 arrivals		-0.589		
1950-1954 arrivals		-0.547		
1955-1959 arrivals		-0.540		
1960-1964 arrivals		-0.487		
1965-1969 arrivals		-0.367		
1970-1974 arrivals		-0.390		
1975-1979 arrivals		-0.417		
1980-1984 arrivals		-0.409		
1985-1989 arrivals		-0.406		
1990-1994 arrivals		-0.440		
1995-1999 arrivals		-0.305		
2000-2004 arrivals		-0.300		
2005 arrivals		-0.369		
By Continent of Nationality				
EU (except A10)			-0.409	
A10			-0.525	
Europe (except EU)			-0.388	
Asia and Middle East			-0.521	
North America			-0.160	
Central and South America			-0.521	
Africa			-0.511	
Australasia and Oceania			-0.314	
Unknown			-0.567	
Immigrant-Native Earnings Gap After:				
Selected continent or cohort		*1945-1949*	*North America*	
10 years	-0.281	-0.418	-0.014	
20 years	-0.042	-0.175	0.204	
30 years	0.104	-0.015	0.335	
40 years	0.156	0.062	0.379	
Selected continent or cohort		*2000-2004*	*Africa*	
10 years		-0.129	-0.365	
20 years		0.114	-0.147	
30 years		0.274	-0.016	
40 years		0.351	0.028	
Immigrant Earnings Growth Rate After:				
10 years	0.126	0.176	0.112	
20 years	0.233	0.328	0.199	
30 years	0.322	0.455	0.263	
40 years	0.392	0.557	0.302	
Immigrant-Native Earnings Convergence Rate After:				
10 years	0.333	0.326	0.305	
20 years	0.573	0.569	0.522	
30 years	0.718	0.729	0.653	
40 years	0.770	0.806	0.698	

Table 2 - Earnings Gap, Earnings Growth and Earnings Convergence

(I) Estimates derived from coefficients in Table A3. All estimates are significant at 5% (see Table A3 for standard errors). See text fo

Table A1 - Immigrant-Native Earnings Gap by Year

Tax-Year	At Entry				After 1 year		After 10 years		After 20 years		After 30 years	
	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors
1978	-0.708	0.086	-0.459	0.081	-0.041	0.069	-0.090	0.089	0.017	0.062	-0.328	0.071
1979	-0.808	0.086	-0.467	0.078	-0.130	0.060	-0.157	0.102	0.030	0.058	-0.207	0.062
1980	-0.813	0.104	-0.466	0.096	-0.032	0.063	-0.079	0.078	0.031	0.057	-0.190	0.072
1981	-0.972	0.131	-0.555	0.096	0.061	0.081	-0.154	0.077	0.116	0.055	0.013	0.069
1982	-0.831	0.150	-0.429	0.128	-0.006	0.092	-0.046	0.125	0.138	0.066	0.062	0.079
1983	-1.168	0.129	-0.671	0.118	0.005	0.096	0.062	0.056	0.072	0.071	0.043	0.092
1984	-0.967	0.114	-0.502	0.095	-0.135	0.076	-0.049	0.028	0.237	0.070	0.059	0.076
1985	-0.798	0.097	-0.429	0.074	-0.023	0.081	0.091	0.034	0.293	0.065	0.108	0.073
1986	-0.690	0.089	-0.309	0.080	-0.202	0.065	0.120	0.040	0.230	0.067	0.134	0.064
1987	-0.930	0.083	-0.474	0.074	0.096	0.066	0.044	0.043	0.239	0.071	0.149	0.081
1988	-0.921	0.081	-0.549	0.070	-0.035	0.056	0.167	0.049	0.281	0.072	0.187	0.080
1989	-0.723	0.071	-0.512	0.064	-0.058	0.053	0.021	0.049	0.201	0.079	0.139	0.084
1990	-0.927	0.074	-0.590	0.064	-0.049	0.052	0.003	0.052	0.223	0.096	0.156	0.069
1991	-0.790	0.092	-0.408	0.078	-0.109	0.052	0.043	0.049	0.291	0.093	0.254	0.071
1992	-0.808	0.092	-0.626	0.071	-0.096	0.057	0.029	0.048	0.332	0.161	0.239	0.082
1993	-1.033	0.096	-0.514	0.076	-0.018	0.065	0.015	0.047	0.286	0.076	0.212	0.080
1994	-0.960	0.079	-0.608	0.061	-0.072	0.058	0.033	0.043	0.013	0.036	0.385	0.091
1995	-1.036	0.073	-0.640	0.062	-0.134	0.051	0.013	0.044	0.218	0.047	0.344	0.089
1996	-0.643	0.066	-0.286	0.059	-0.028	0.046	0.086	0.045	0.156	0.049	0.491	0.090
1997	-0.627	0.101	-0.353	0.085	0.062	0.052	0.014	0.044	0.140	0.052	0.224	0.084
1998	-0.924	0.061	-0.594	0.047	-0.103	0.061	0.016	0.039	0.210	0.058	0.544	0.093
1999	-0.972	0.055	-0.173	0.045	0.134	0.034	0.088	0.039	0.158	0.061	0.255	0.112
2000	-0.839	0.051	-0.112	0.041	0.133	0.032	0.134	0.042	0.117	0.059	0.270	0.081
2001	-0.807	0.051	-0.101	0.041	0.203	0.028	0.128	0.043	0.240	0.058	0.308	0.081
2002	-0.982	0.043	-0.192	0.035	0.150	0.028	0.157	0.041	0.155	0.056	0.352	0.161
2003	-1.027	0.037	-0.206	0.031	0.153	0.024	0.133	0.040	0.198	0.052	0.403	0.076
2004	-1.231	0.032	-0.392	0.026	0.020	0.024	0.170	0.038	0.141	0.051	0.091	0.040
2005	-1.135	0.028	-0.289	0.023	0.062	0.018	0.187	0.036	0.181	0.046	0.153	0.052
2006	0.000	0.000	-	-	0.096	0.020	0.135	0.037	0.120	0.051	0.217	0.056
Number of employed wee...	No		Yes		Yes		Yes		Yes		Yes	
Number of individuals												
Natives	387760		387760		387760		387760		387760		387760	
Immigrants	9949		9949		17625		11658		7060		4354	

(1) The sample data used includes those aged 25 to 64 earning between £100 and £1000000 in any one tax year who are observed at least twice, hence, by construction, all t
   entering in 2006 were dropped because they were only observed once (it excludes the self-employed). It includes immigrants arriving from 1945 onwards. See text for de
(2) The number of immigrants in the year after entry is higher than the number in the entry year because the entry date is only recorded when the immigrant applies for a NIN
   depends on her individual circumstances and might not be immediately upon arrival. See text for details.

Table A2 - Immigrant-Native Earnings Gap by Year (tracked immigrants)							
Year	After 1 year		After 10 years		After 20 years		After 30 years	
	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors
1978	0.081	0.149	0.238	0.097	0.175	0.078	-0.328	0.071
1979	-	-	0.155	0.187	0.176	0.074	-0.207	0.062
1980	-	-	0.135	0.121	0.225	0.076	-0.190	0.072
1981	-	-	-0.078	0.106	0.285	0.062	0.013	0.069
1982	-	-	-0.021	0.145	0.309	0.088	0.062	0.079
1983	-	-	0.182	0.087	0.150	0.094	0.043	0.092
1984	-	-	0.008	0.039	0.290	0.082	0.059	0.076
1985	-	-	0.084	0.043	0.356	0.078	0.108	0.073
1986	-	-	0.089	0.056	0.289	0.081	0.134	0.064
1987	-	-	-	-	0.449	0.081	0.149	0.081
1988	-	-	-	-	0.342	0.087	0.187	0.080
1989	-	-	-	-	0.291	0.094	0.139	0.084
1990	-	-	-	-	0.305	0.115	0.156	0.069
1991	-	-	-	-	0.259	0.118	0.254	0.071
1992	-	-	-	-	0.334	0.243	0.239	0.082
1993	-	-	-	-	0.443	0.081	0.212	0.080
1994	-	-	-	-	0.142	0.043	0.385	0.091
1995	-	-	-	-	0.235	0.054	0.344	0.089
1996	-	-	-	-	0.261	0.060	0.491	0.090
1997	-	-	-	-	-	-	0.224	0.084
1998	-	-	-	-	-	-	0.544	0.093
1999	-	-	-	-	-	-	0.255	0.112
2000	-	-	-	-	-	-	0.270	0.081
2001	-	-	-	-	-	-	0.308	0.081
2002	-	-	-	-	-	-	0.352	0.161
2003	-	-	-	-	-	-	0.403	0.076
2004	-	-	-	-	-	-	0.091	0.040
2005	-	-	-	-	-	-	0.153	0.052
2006	-	-	-	-	-	-	0.217	0.056
Number of employed we	Yes		Yes		Yes		Yes	
Number of individuals								
Natives	146246		223960		314807		387760	
Immigrants	179		2131		3659		4354	

(I) Notes as in Table A1, except that the sample now tracks immigrants who stayed for 30 years or longer.

Table A3 - Earnings Regression

Variable	coefficient	s.errors	coefficient	s.errors	coefficient	s.errors	coefficient male	s.errors	coefficient female	s.errors
Intercept	6.062	0.013	6.067	0.013	6.040	0.013	5.983	0.018	7.005	0.020
Immigrant (=1)	-0.462	0.051					-0.449	0.074	-0.573	0.068
1945-1949 arrivals			-0.589	0.053						
1950-1954 arrivals			-0.547	0.056						
1955-1959 arrivals			-0.540	0.056						
1960-1964 arrivals			-0.487	0.055						
1965-1969 arrivals			-0.367	0.054						
1970-1974 arrivals			-0.390	0.054						
1975-1979 arrivals			-0.417	0.052						
1980-1984 arrivals			-0.409	0.053						
1985-1989 arrivals			-0.406	0.053						
1990-1994 arrivals			-0.440	0.052						
1995-1999 arrivals			-0.305	0.053						
2000-2004 arrivals			-0.300	0.054						
2005 arrivals			-0.369	0.056						
EU (except A10)					-0.409	0.052				
A10					-0.525	0.054				
Europe (except EU)					-0.388	0.056				
Asia and Middle East					-0.521	0.053				
North America					-0.160	0.054				
Central and South America					-0.521	0.055				
Africa					-0.511	0.054				
Australasia and Oceania					-0.314	0.053				
Unknown					-0.567	0.051				
Years since immigration	0.014	0.001	0.019	0.001	0.012	0.001	0.022	0.001	0.008	0.001
Years since immigration squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sex (male=1) x Immigrant (=1)	-0.160	0.007	-0.164	0.007	-0.166	0.007				
Age x Immigrant (=1)	0.024	0.003	0.018	0.003	0.022	0.003	0.008	0.004	0.031	0.004
Age squared x Immigrant (=1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of employed weeks x Immigrant	0.001	0.000	0.001	0.000	0.001	0.000	0.003	0.000	0.000	0.000
Number of jobs x Immigrant (=1)	-0.017	0.005	-0.020	0.005	-0.015	0.005	-0.007	0.007	-0.018	0.006
Sex (male=1)	0.604	0.002	0.604	0.002	0.604	0.002				
Age	0.047	0.000	0.047	0.000	0.047	0.000	0.088	0.001	-0.007	0.001
Age squared	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
Number of employed weeks	0.040	0.000	0.040	0.000	0.040	0.000	0.037	0.000	0.042	0.000
Number of jobs	0.349	0.003	0.349	0.003	0.349	0.003	0.319	0.004	0.368	0.004
R-squared	0.49		0.49		0.50		0.38		0.49	
Sample size	5366162		5366162		5366162		3032734		2333428	
Number of individuals	433069		433069		433069		234982		198087	

(1) The sample data used includes those aged 25 to 64 earning between £100 and £1000000 in any one tax year who are observed at least twice (it excludes the self-employed). It includes immigrants arriving from 1945 onwards. See text for details.
(2) All models include area fixed effects (49 county dummies) and time fixed effects (29 tax-year dummies). See text for details.
(3) All models are corrected for intragroup correlation, as standard errors are assumed independent across groups of individuals but not within groups (i.e. errors are not assumed independent for a particular individual over time). See text for details.
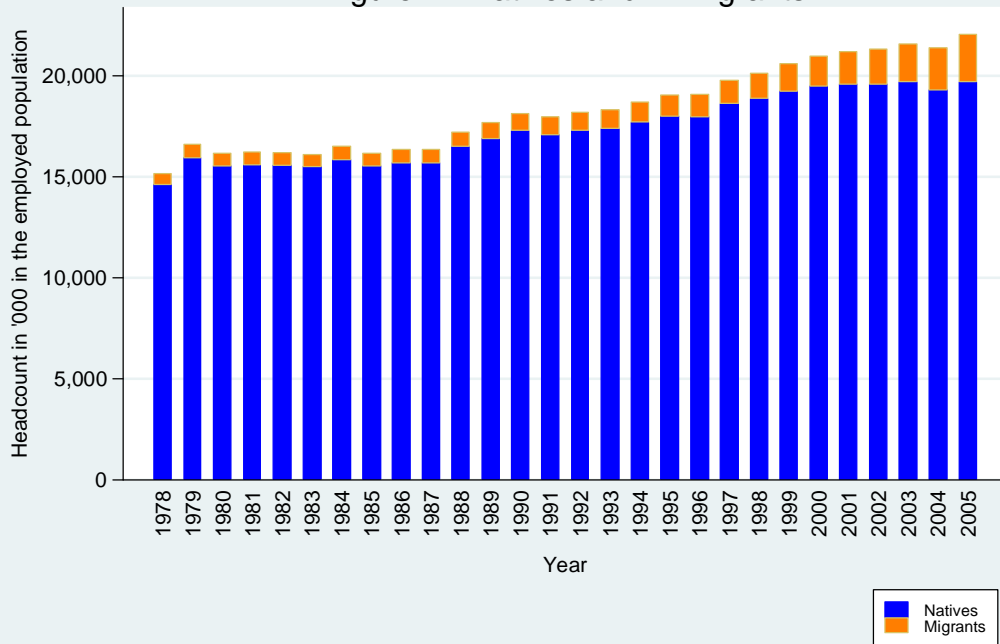
Figure 1 - Natives and Immigrants

Source: Lifetime Labour Market Database

Figure 2 - Immigrants by Continent of Nationality

Source: Lifetime Labour Market Database
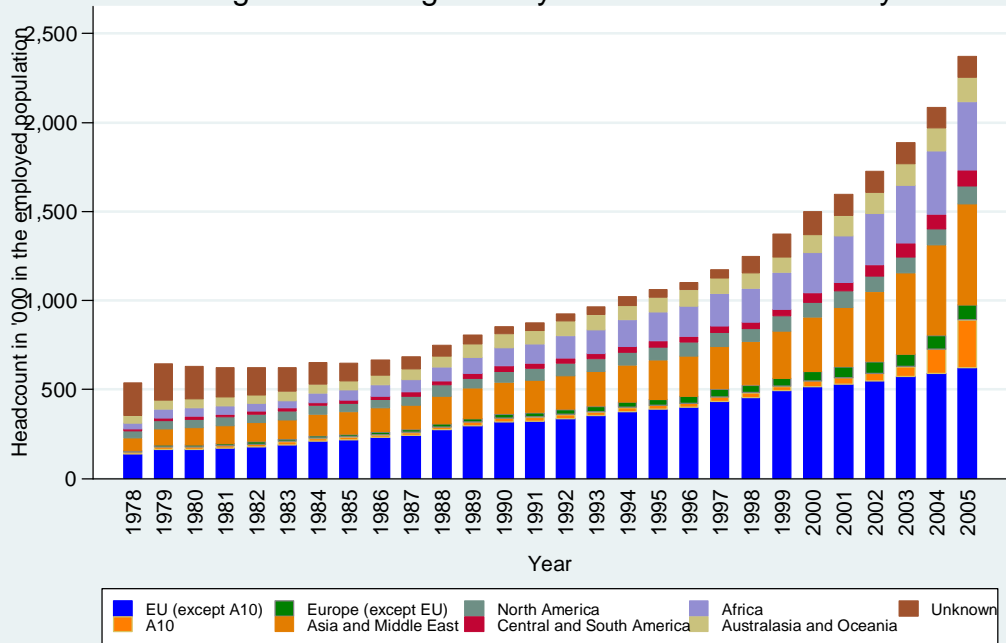
Figure 3 - Immigrants' Inflow by Continent of Nationality

Source: Lifetime Labour Market Database
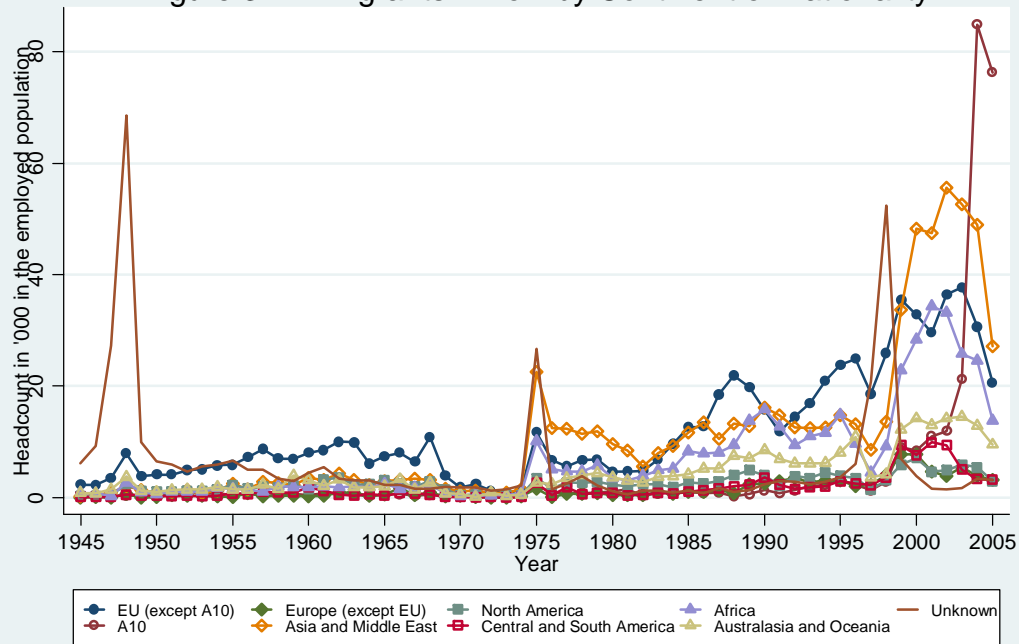
Figure 4 - Immigrant-Native Earnings Gap

Source: Estimates derived from coefficients in Table A1 (see text for details)
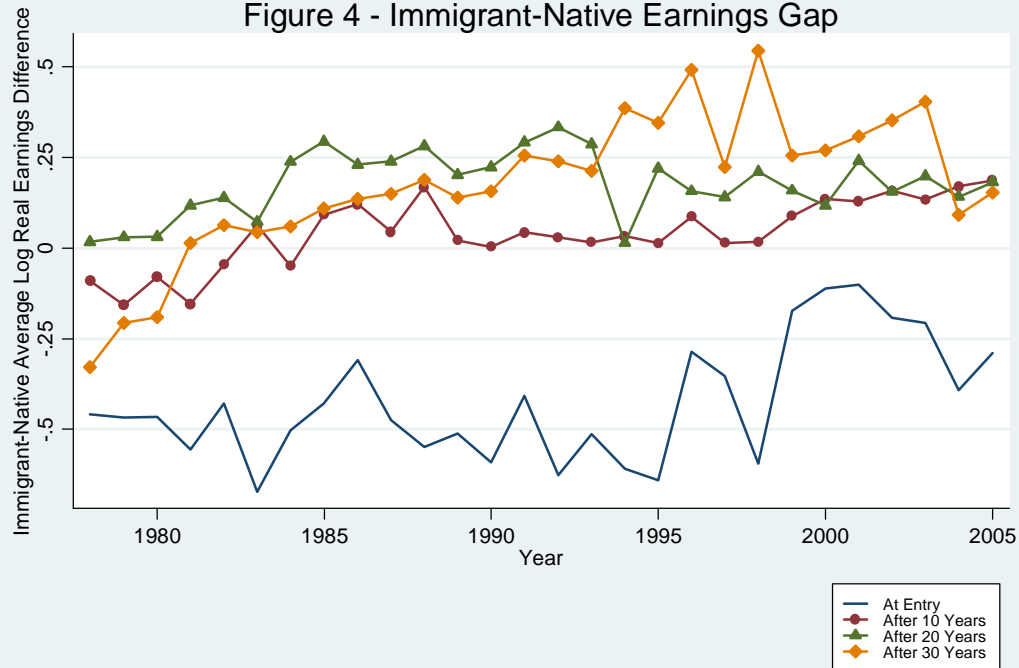
Figure 5 - Immigrant-Native Earnings Gap After 20 Years of Arrival

Source: Estimates derived from coefficients in Table A2 (see text for details)
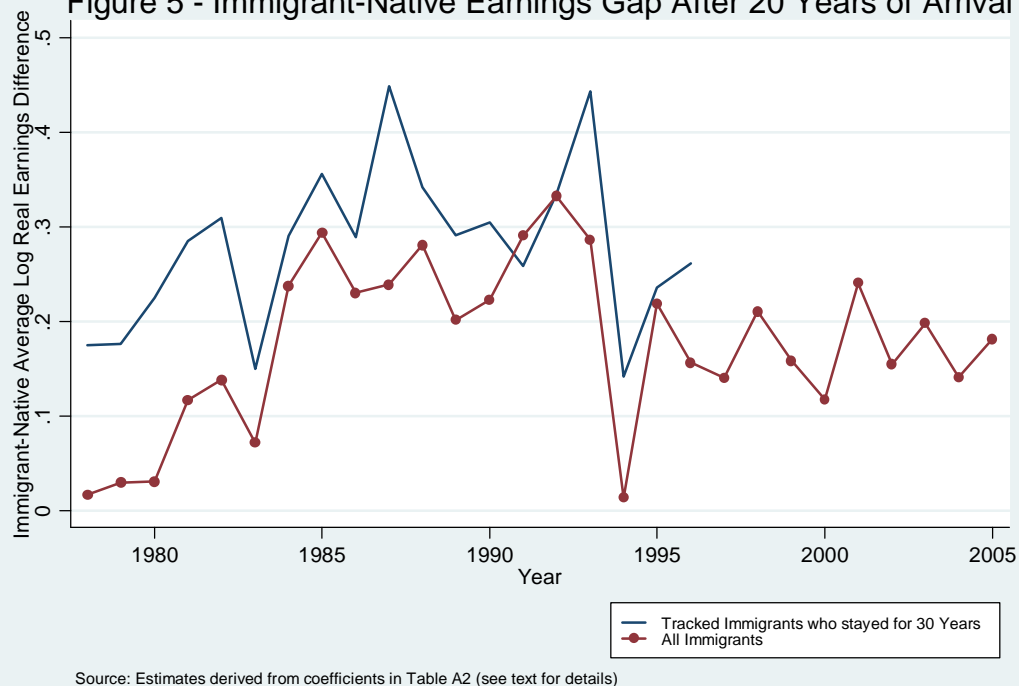
Figure 6 - Immigrant-Native Earnings Gap by Cohort of Arrival

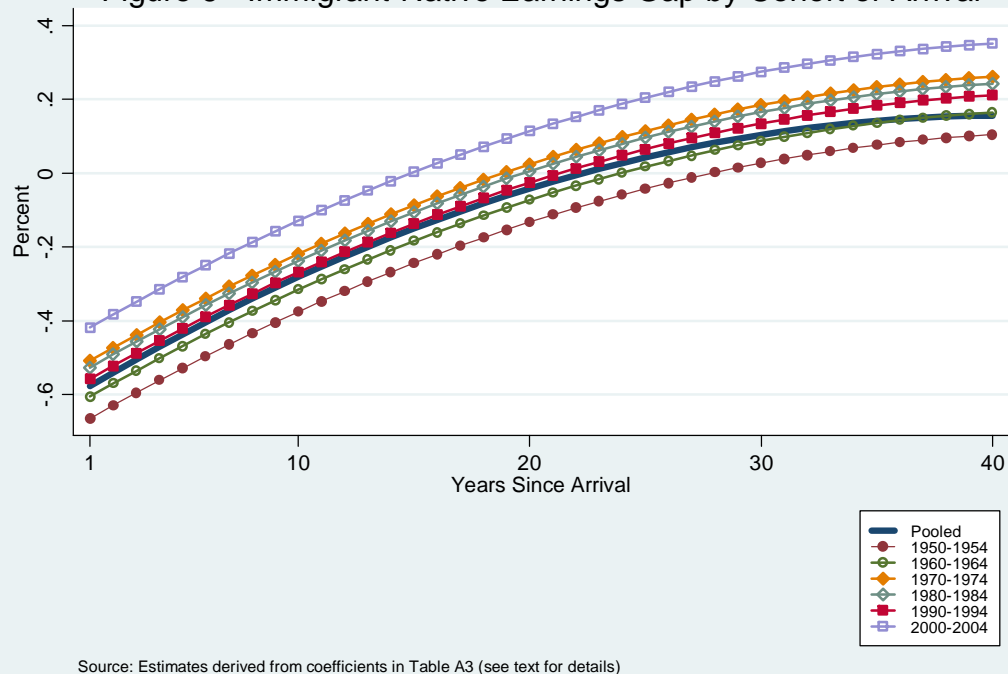Source: Estimates derived from coefficients in Table A3 (see text for details)

Figure 7 - Immigrant-Native Earnings Gap by Continent of Origin

Source: Estimates derived from coefficients in Table A3 (see text for details)