Leading education and social research
Institute of Education
University of London

Department of Quantitative Social Science

Over-education among A8 migrants in the UK

Stuart Campbell

DoQSS Working Paper No. 13-09
May 2013

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

DoQSS Workings Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Department of Quantitative Social Science, Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, UK

Over-education among A8 migrants in the UK

Stuart Campbell[1]

Abstract

I present new evidence on the incidence and wage associations of over-education among migrants to the UK from the 'A8' EU accession countries of Central and Eastern Europe from 2004-2011. Using the Labour Force Survey, I employ a novel strategy to maximise the number of migrants drawn from the dataset over the period of interest, creating a survey sample of A8 migrants of unprecedented size. I also use a new method of classifying education attained outside the UK, which takes account of different European education systems. I find that A8 migrants face a substantially higher risk of over-education in the UK than other recent EU migrants, and that this additional risk remains after taking account of observed characteristics. I argue that this result is driven by unobserved differences between the groups, arising from distinct self-selection processes associated with the institutional context of the EU accession. I also find that in non-graduate occupations, the wage penalties faced by A8 migrants in the UK are of such strength that even the over-educated are paid less than matched UK nationals. Moreover, A8 migrants are concentrated in a particular sub-group of occupations, where higher wages are not available for the over-educated.

JEL classification: J24; J61; J62; F22

Keywords: Immigration, educational mismatch, labour market, skill recognition

[1] Department of Quantitative Social Science, Institute of Education, University of London (scampbell@ioe.ac.uk)

Acknowledgements:

This work is supported by the Economic and Social Research Council, via the Bloomsbury Doctoral Training Centre. Data are provided by the UK Data Service. I am grateful for conscientious supervision from John Micklewright and Lucinda Platt. I also received helpful comments on an earlier draft from Anna Vignoles and Charlie Owen, and at presentations of various iterations of this work, including at the 2013 NORFACE migration conference, the 2012 OECD/CEPII Immigration Workshop (particularly from Isabelle Chort), the 2012 ESDS LFS User Meeting, and at seminars in the Department of Quantitative Social Science at the Institute of Education, University of London.

1. Introduction

The highly educated migrant earning a modest living as a cleaner, shop assistant, or factory worker is a popular character in public discussions of immigration in the rich world, and there is some evidence to suggest that the phenomenon of migrant 'over-education' has more than a merely anecdotal existence. Where it does exist, such mismatch between occupation and educational background potentially represents a waste of individual potential for the migrant affected, as well as a failure for the host country to capture the full economic gains from immigration. In the United Kingdom, where the leaders of all three major political parties have expressed concern that low-skilled native workers have suffered from unfair competition arising from increased immigration in recent years,[2] the perception that over-qualified migrants are displacing low-skilled natives may also be damaging for social cohesion, and for public support of moderation in immigration policy-making.

A large empirical literature suggests that over-education is associated with decreased job satisfaction, higher labour market turnover, and reduced earnings potential.[3] For migrants, it may also contribute to a decision to move on to another country, or indeed to move home. Recent migrants to the UK from the EU accession countries of Central and Eastern Europe (the 'Accession 8', or 'A8' countries), namely the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, and Slovenia, have developed a strong reputation for being over-educated, but no quantitative study has yet investigated the incidence and implications of over-education in this group. This represents an important omission from the literature, given both the grand scale of this wave of migration, which some believe to have been the largest to the UK in history (Salt and Rees, 2006), and its unique character, which appears to have been more temporary and recurrent than that observed in other migrant groups in the UK (see Eade *et al*., 2007, 33-34). The purpose of this paper is to establish the prevalence of over-education among A8 migrants in the UK, and to investigate any potential wage implications.

The main comparison group I use in this study is EU15 migrants, who come to the UK from countries which were already EU members in 2004 (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain and Sweden).[4] Migrants from these countries have been coming to work in the UK over a longer time period, but to allow comparison with the recent A8 migrants, I only consider those arriving in 2004-2011. UK nationals act as an alternative comparison group.

[2] See Cameron (2011); Clegg (2013); Miliband (2013). Empirical support for such claims is mixed. Dustmann *et al.* (2013) find that immigration depresses wages at the lower end of the UK wage distribution, but the evidence on employment effects is weaker (Dustmann *et al.*, 2005, find no well determined impact of immigration on native employment, while the Migration Advisory Committee, 2012: 10, find a 'tentative negative association' under certain macroeconomic conditions).

[3] Various empirical studies on the implications of over-education are cited in Allen and van der Velden (2001: 434).

[4] The 'EU15' designation usually includes the UK, but UK nationals are treated as a separate group here. Malta and Cyprus joined the EU at the same time as the A8 countries, but under quite different institutional circumstances, and nationals of these countries are therefore excluded from the analysis. The same applies to migrants from Romania and Bulgaria (the 'A2' countries), who have been part of the EU since 2007, but will not enjoy equal labour market access in the UK until January 2014.

The definition of 'over-education' I use in this study is based on the distribution of educational attainment within a given occupation. Having established a standard level of education within each occupation, using an internationally comparable measure of educational attainment, I class individuals as 'matched', or 'over-educated', depending on how their own level of educational attainment compares to this standard level. I then compare the prevalence of over-education in different migrant groups. I take account of differences in the observable characteristics of the migrant groups using a probability model, and I assess the wage associations of over-education using a variant on the standard human capital earnings function.

The data I use here come from the Labour Force Survey (LFS) between 2004 and 2011. Previous empirical work on A8 migrants in the UK has been hindered by the small survey sample sizes possible, even with large datasets such as the LFS, and by the difficulty of estimating returns to education attained in different European education systems. I use a novel strategy to extract information on the maximum number of individuals possible from the LFS, which results in a cross-sectional sample substantially larger than any used in previous survey-based studies of A8 migrants. I also use an improved method of classifying the past educational attainment of A8 migrants, which takes account of the differences between European education systems.

A recent review of the scholarly literature examining over-education among migrants in general is provided by Piracha and Vadean (2012). Almost all studies in this area have suggested that migrants suffer a higher propensity to be over-educated than the native population, and that migrants receive lower returns to surplus education than natives. However, I argue that the institutional context of the EU accession attracted and retained migrants from the A8 countries with a unique vulnerability to over-education, and my findings are consistent with this hypothesis. I also find that mismatch is associated with wage differences, but that in this case they are dominated by strong migrant wage penalties. Further, I note that A8 migrants are concentrated in a particular sub-group of occupations, where higher wages are not available for the over-educated.

In Section 2, I present background information on A8 migrants in the UK and review some of the existing evidence relating to this group; in Section 3, I discuss the dataset and the definitions I use in this study; in Section 4, I examine the prevalence of over-education and the characteristics with which it is associated; in Section 5, I examine the wage associations of this over-education, and in Section 6, I conclude, and make a brief comment on policy.

2. Background

2.1 Who are the A8 migrants?

The 'A8' countries are the eight former communist countries of Central and Eastern Europe that joined the EU on May 1st 2004. EU citizens had previously been allowed almost free access to the labour markets of the other member states, but fears about mass immigration from the poorer A8 countries led to the establishment of 'transitional arrangements' in the richer EU15. For most governments of the EU15, these transitional arrangements involved placing substantial barriers to the employment of A8 migrants,[5] but the governments of the UK, Ireland and Sweden allowed more or less free access to their labour markets. The UK has by far the largest labour market of these countries, and, although the exact numbers are contentious, perhaps 1.5 million A8 migrants came to the UK to work for some period of time in the first five years following accession in 2004 (Sumption and Somerville, 2010: 5).

The transitional arrangements in the UK from May 2004 until the end of April 2011 placed some restrictions on access to welfare benefits for A8 migrants in their first year in the country, as well as requiring initial registration on a 'Worker Registration Scheme' (WRS) in order to take up employment. The demographic information collected in the WRS is thoroughly summarised in Drinkwater *et al*. (2009: 166-167) and in Blanchflower and Shadforth (2009: F145-F146). Broadly, it indicates that most A8 workers in the UK are aged between 18 and 34, and only a small number arrive with dependent children. A8 nationals also appear to be more evenly distributed around the country than other migrant groups, which tend to be clustered in large metropolitan areas (Drinkwater *et al*., 2009: 167).

The speed and scale of A8 migration drew scholarly attention, focussing, for example, on its impact on the domestic labour market (Portes and French, 2005; Gilpin *et al*., 2006; Green *et al.*, 2007; Lemos and Portes, 2008), the housing market (Robinson, 2007; Pemberton, 2009), its fiscal effects (Dustmann *et al*. 2010), and its impact on the crime rate (Bell *et al*., 2010). The geographical distribution of the early A8 migrants has also been addressed (Coombes *et al.*, 2007). Other authors have considered the labour market performance of A8 migrants, and the returns to education available for these workers in the UK. For example, Clark and Drinkwater (2008) found A8 migrants worked relatively long hours for relatively low wages, and had particularly low returns to education. Drinkwater *et al*. (2009) also found low returns to education among A8 migrants.

[5] Restrictions on the rights of A8 migrants to work in all of the other EU15 countries had ended by May 1st 2011, with the end of the period of 'Transitional Arrangements'.

Over-education among this group of migrants in the UK has been considered in several qualitative studies, often focussing on one or more of the constituent A8 nationalities.[6] Parutis (2011) explicitly addresses the question of why so many Polish and Lithuanian migrants in the UK appear to be working below their level of qualification, using in-depth interviews. She argues that often the motivation for migration does not centre on wage benefits, and that learning English, and the experience of living abroad, can play an important role. Similarly, Anderson *et al*., (2006) find over-qualified interviewees explaining their situation both in terms of such non-wage benefits, and in terms of a conscious economic trade-off, as a low wage in the UK may be relatively high when compared with wages in the country of origin. There is also qualitative evidence of discrimination against A8 migrants causing over-education, or at least a lack of recognition of qualifications attained in A8 countries. For example, Cook *et al.* (2011a: 61) note that more highly qualified workers expressed frustration that imported qualifications and experience were not recognised by employers in the UK. Such interview evidence is very valuable, particularly in assessing individual perspectives on the causes and consequences of educational mismatch, but it is difficult to judge the reliability of these subjective accounts of over-education, or indeed their national significance.

This previous research has therefore indicated that the reputation for over-education among A8 migrants has some empirical basis. However, no study has yet produced quantitative evidence that addresses over-education in this group explicitly, or that examines the association between over-education and wages. As such, this paper should be a useful addition to the existing research.

[6] On the possible consequences of over-education, aside from labour market implications, over-education among Polish migrants in Scotland has also been cited in the public health literature as a major contributor to stress (Weisharr, 2008: 1253).

2.2 How many A8 migrants are in the UK?

It is difficult to estimate the number of A8 migrants who have come to work in the UK with precision: no comprehensive official records are kept of people entering or leaving the country, and researchers must therefore rely on survey evidence, which often struggles to capture recent migrant populations adequately (see the discussion in Section 3.1 below), or on domestic administrative data, which often lacks detail, and is not always comprehensive. Around one million workers from the A8 countries registered on the 'Worker Registration Scheme' (WRS) in the UK in the first five years after accession, and, taking account of the many workers who did not register on the scheme, it has been inferred that around 1.5 million A8 workers came to the UK in total, though much of this migration has been temporary (Sumption and Somerville, 2010: 9).

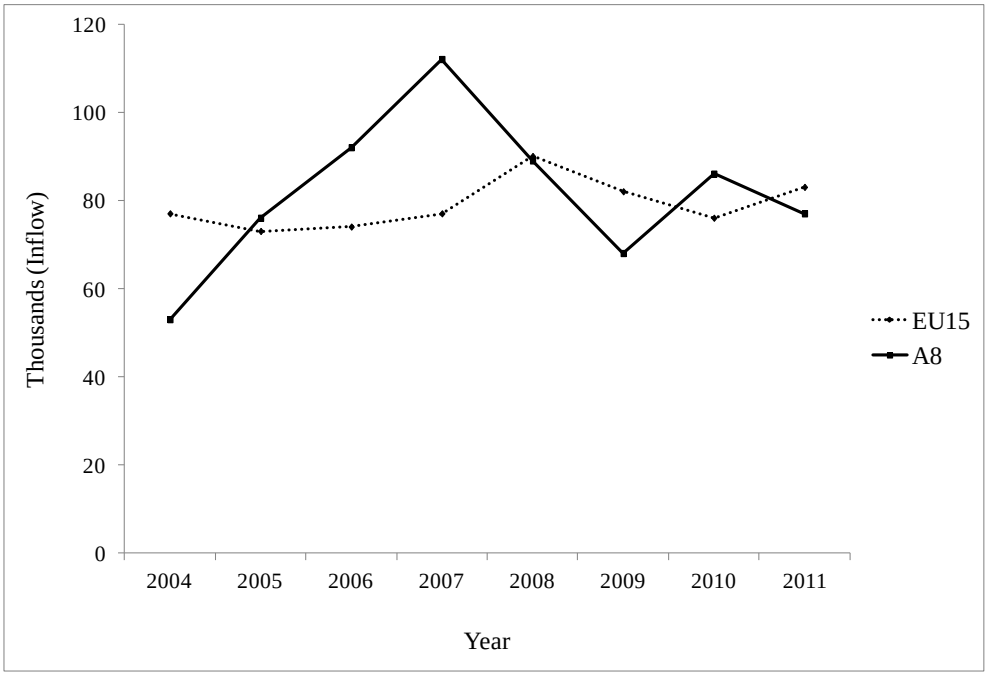
Leaving aside these problems of accurate measurement, it is clear that at any point in time, A8 migrants make up a small proportion of the UK working age population, which is now around 38.5 million people (ONS, 2012a). One consequence of this is that even a large government survey such as the LFS can capture only a relatively small number of A8 migrants, and this creates problems for statistical inference. I employ a novel strategy to increase sample size, discussed below in Section 3.3.

Long-Term International Migration (LTIM) data, which are based on the International Passenger Survey (IPS), suggest total net migration from the A8 countries over 2004 to 2011 of only around 393 thousand (compared to 354 thousand for EU15 migrants), but the IPS uses an interpretation of the United Nations definition of a long-term international migrant, which specifies that a person must stay in the country for at least a year in order to be properly considered as such. As the IPS is conducted at the point of arrival in the UK, migrants are asked about their 'intended' length of stay, and included in the survey if this is over one year.[7] This definition excludes many A8 and EU15 migrants with short time horizons in the UK, including those who end up staying for longer than a year, for there is often a large discrepancy between intended and actual length of stay in the country (Clark and Drinkwater, 2008: 504n). In order to give a sense of trends in long-term migration from the A8 countries and the rest of the EU, at least in the limited sense of 'intended' long-term migration, Figure 1 compares the total inflow and outflow of A8 and EU15 migrants recorded in the LTIM data, over 2004-2011.

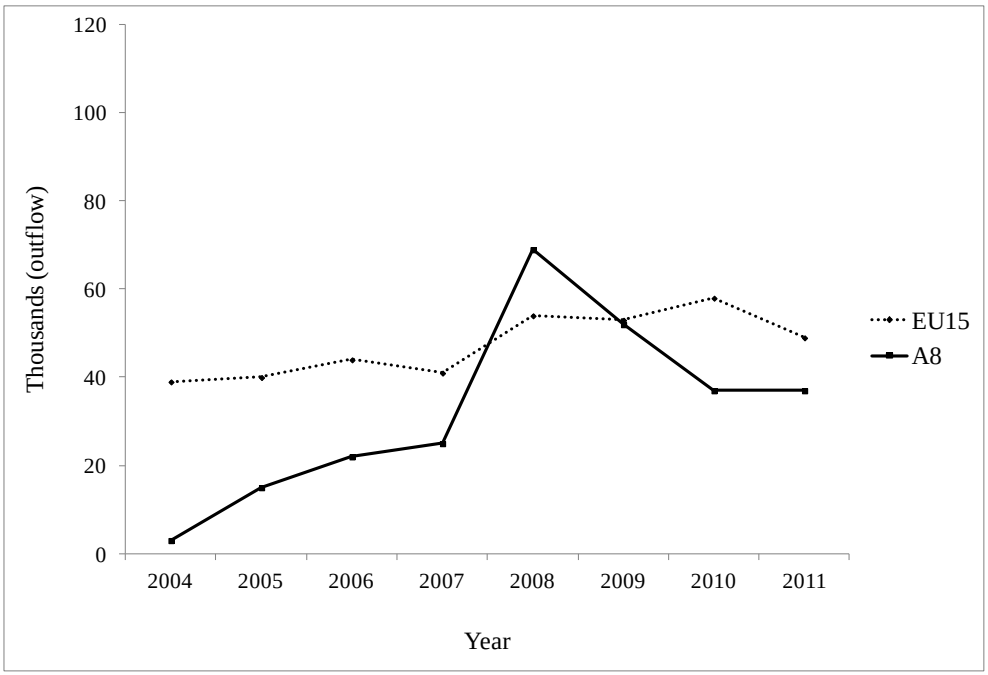
[7] 'Outflow' estimates are collected at the point of departure in the UK, and migrants are identified based on their actual length of their stay.

Figure 1: Total flows of Long-term A8 and EU15 migrants to the UK, 2004-2011

(a) Inflows ('Intended' Long-term migrants)



(b) Outflows ('Actual' long-term migrants)



Source: LTIM time-series data, ONS (2013a).

The inflow and outflow of EU15 migrants appears to be relatively constant during the transitional arrangements, with an increase in the rate of outflow over the second-half of the period. In contrast, the inflow of A8 migrants rises sharply up to 2007, and then falls sharply until 2009, before starting to rise again in 2010, and falling slightly in 2011. The outflow of A8 migrants rises more slowly until 2007, before increasing sharply in 2008, and then falling back again over 2009-2010. These patterns suggest that the flow of A8 migrants with long-term intentions is less constant then that of EU15 migrants with long-term intentions. Indeed, with the fluctuations in the UK macro-economy since 2008, it may be that the A8 migrant group is simply more responsive to macroeconomic conditions. This would be consistent with the evidence that patterns of A8 migration are more fluid than those of other migrant groups, an idea that I will refer to again in the next section.

2.3 Why are A8 migrants different?

I argue here that A8 migrants face a higher risk of over-education than recent EU15 migrants because of unobserved differences in their labour market characteristics, and possibly also because they experience a higher degree of labour market discrimination in the UK. These unobserved differences in characteristics include more heterogeneous motivations, more uncertain time-horizons, and lower reservation wages. Such qualities reflect a distinct self-selection process associated with the institutional context of the accession. Specifically, wide real wage gaps, wide differences in absolute price levels, and a one year exclusion from government welfare benefits attracted and retained migrants with a unique vulnerability to educational mismatch.

Migrants tend to be positively self-selected for labour market ability and motivation (Chiswick, 1978). However, in general, the larger real wage gaps between the UK and the A8 home countries[8] imply that the inflow of migrants from these countries will be less positively self-selected for such characteristics than will the inflow of migrants from the richer EU15 countries, if transport and other fixed costs of migration are reasonably similar.[9] In other words, migration to the UK will be a profitable enterprise for many people from the A8 countries whether their labour market ability and motivation is high or low, while only the most able and most motivated workers in EU15 countries will gain from migrating. Indeed, the heterogeneity of the motivations for migration cited by people from A8 countries is one of the features of the qualitative literature in this area (Anderson *et al.*, 2006; Cook *et al.*, 2011a; Parutis, 2011), and this may reflect the more marginal role that labour market motivations play for some such migrants.

[8] I do not quantify the wage gaps between the A8 and EU15 countries here, but Eurostat (2013) provides detailed maps of GDP per capita within the EU over the period of the transitional arrangements.
[9] This is a straightforward implication of the human capital approach to migration in Chiswick (1978). See Chiswick (2001) for a discussion of migration costs and positive selection.

The larger real wage gaps between the A8 countries and the UK may also foster a higher occurrence of 'temporary' or 'circular' migration,[10] as extended periods of country-specific human capital adaptation are not required in order to make migration profitable. Time-horizons in the UK therefore may be less certain for many A8 migrants. Indeed, one of the stylised facts emerging from the empirical work on A8 migrants so far is that they are an unusually fluid group of migrants, with many people coming to the UK and working for only a short time before moving elsewhere, or moving back and forth between their home country and the UK over a longer period (see Eade *et al.*, 2007: 33-34). Engbersen *et al.* (2010) have described the movement of A8 workers around the EU as 'liquid migration', with workers 'trying their luck' in different European labour markets before settling or moving on.[11] If those A8 migrants with greater labour market ability leave the UK permanently after achieving some pre-determined level of target savings, then the relative degree of positive self-selection for labour market ability and motivation will be further reduced in the A8 migrant stock in the UK.

At the same time, in terms of generic skills associated with securing employment quickly, the inflow of A8 migrants will be strongly positively self-selected, since A8 nationals are excluded from government benefits during their first year in the UK, and the gap between the absolute price levels in the UK and the A8 countries makes any period out of work particularly costly. Thus, in order to secure positive returns to migration, many will have to be able to find work fast, and with minimum expenditure.[12] In a job-search framework, such workers can be said to have a lower 'reservation wage'. EU15 migrants are not excluded from government benefits, and do not face an equivalent price gap, and so job-search is not so constrained. If those A8 migrants who are unable to secure or maintain employment leave the UK permanently, this will leave the remaining stock of such migrants even more strongly selected for these characteristics. The lower reservation wages among many A8 migrants are reflected in exceptionally high employment rates, which averaged 81.1% over the years of the transitional arrangements, compared to 72.4% for EU15 nationals and 72.8% for UK nationals.[13] They are also reflected in the unusually high geographic mobility of this migrant group, as workers travel to areas of labour-shortage, rather than clustering in large metropolitan areas (Coombes *et al.*, 2007).

A8 migrants may also face higher levels of employer discrimination. Little work has been done so far in the quantitative literature on discrimination specifically against A8 migrants, but, for example, there is qualitative evidence of general hostility towards A8 migrants from the host population in the UK (see Cook *et al.*, 2011a: 61-62; Cook *et al.*, 2011b: 736; Spencer *et al*. 2007: 66-69; Anderson *et al.*, 2007: 15; and a review of newspaper coverage of Polish migrants in Fomina and Frelak, 2008), and it is plausible that such

[10] For a concise summary of different forms of migration, see Dustmann and Weiss (2007: 237-238).
[11] Eade *et al*. (2007: 34) have referred to the 'intentional unpredictability' of such migrants.
[12] The high relative cost of job-search for A8 migrants may be reflected in their more intensive use of social contacts when looking for work. Sumption (2009) presents evidence of this for Polish migrants, and Battu *et al.* (2011) note similar patterns for the A8 group as a whole.
[13] These are my calculations, from the tables in ONS (2013b).

discrimination could be affecting occupational outcomes for A8 migrants in a way that it does not affect outcomes for EU15 migrants. And, as noted above, some more highly qualified A8 workers have expressed frustration that imported qualifications and experience are not recognised by employers in the UK (Cook *et al.*, 2011a: 61), which may reflect a more subtle form of discrimination, targeted against non-native qualifications and experience, rather than at the individuals that hold them.

I assume that this combination of different motivations, different time horizons, lower reservation wages, and potentially also different levels of labour market discrimination all contribute to a higher risk of over-education for A8 migrants compared to those from EU15 countries. Of course, the distinct geographical, occupational and demographic characteristics of the two groups may also be part of the explanation, but I can account for these characteristics in an econometric analysis, which I will discuss in Section 4.2 below.

In the broader educational mismatch literature, over-educated workers are generally found to earn more than their matched peers within a given occupation. This must partly be because their surplus education proves to be of some productive value to employers (Duncan and Hoffman, 1982). However, migrant groups tend to face wage penalties in the host country independent of educational mismatch, due both to wage discrimination and imperfect human capital portability. For the migrant over-educated, then, wage returns depend on the relative size of the migrant wage penalty and any wage effects associated with over-education. Establishing the relative size of these effects is an empirical exercise, but the existing evidence (e.g. Clark and Drinkwater, 2008; Drinkwater *et al*., 2009) suggests that the wage penalties suffered by A8 migrants in the UK are such that only very strong returns to surplus education would reverse them for the over-educated.

3. Data

3.1 The Labour Force Survey

The LFS is a large sample survey of households in the United Kingdom, which collects a range of demographic and labour market information. It is administered by the Office for National Statistics (ONS). The sample includes around 50,000 responding households each quarter in Great Britain, and around 3,000 responding households in Northern Ireland. As such, it is the largest regular household survey in the UK. The survey employs a rotating panel design, in which addresses are followed quarterly for five successive 'waves', so that around one fifth of the sample is new in each quarter. Addresses to be surveyed are sampled randomly by postcode from the small users' sub-file of the 'Postcode Address File' for Great Britain, and by geographical strata from the 'Valuation List' used for rating purposes in Northern Ireland. Each member of the sampled household is surveyed in person in the first wave, and is then surveyed on the telephone for the four subsequent waves. Interviews may be conducted by proxy if any household member is absent, or if they cannot be surveyed for some other reason, such as poor English language ability. The LFS excludes individuals living in some types of communal establishments, and, until the start of 2008, it also excluded migrants who had arrived in the UK within the preceding 6 months (ONS, 2011: 10).

The LFS follows addresses rather than respondents, and the identity of respondents surveyed may therefore change over the five waves, as current residents leave the household, or as new residents join. Individuals may also be unavailable, or refuse to participate in the survey in any of the five waves. Individual respondents may therefore appear for the first time in waves two, three, four or five, as well as in wave one, and may disappear from the survey intermittently or permanently at any point. I use this feature of the LFS to maximise the sample size in this paper, as I describe in Section 3.3 below.

The LFS has been used extensively to examine the labour market performance of migrants in the UK (for example, Shields and Wheatley-Price, 1998, Blackaby *et al.*, 2005; Dustmann and Fabbri, 2005), including in work focussed specifically on A8 migrants (for example, Clark and Drinkwater, 2008; Jayaweera *et al.*, 2008; Drinkwater *et al*., 2009; Sumption, 2009). There are several reasons to suspect that the LFS under-represents the A8 migrant population. For example, before 2008, all new migrants were excluded by the requirement that they be resident in the UK for six months before being eligible for the survey. Gilpin *et al.* (2006: 11) have also suggested that the exclusion from the LFS of those living in 'communal households' (such as hotels, boarding houses or hostels) may have reduced coverage of A8 migrants in particular.

Aside from these sampling issues in the LFS, a major disadvantage of my using the survey in this application is its lack of regular information on English language ability, which is thought to be a particularly important determinant of labour market success among migrant workers in the UK (see Dustmann and Fabbri, 2003).

Poor English language ability may be reflected in higher rates of response by proxy, but these rates do not differ substantially between A8 migrants, EU15 migrants, and UK nationals. Any household's participation is itself an indicator that at least one available household member has a minimum level of English language ability, which may mean the survey misses the least assimilated migrants.

The potential sampling issues and the lack of data on English language ability together constitute significant weaknesses for conducting research on A8 migrants, but the LFS is of course intended to capture information on the UK labour force generally, and is not a specialist dataset for studying migrants. The survey benefits from collecting the same detailed demographic and work-related information on a large number of respondents from different migrant groups as well as on UK nationals, which represent important points of comparison for A8 migrants in any analysis. Another source of data on A8 migrants is the Worker Registration Scheme (WRS), which covers a much larger sample, but contains only very basic information on demographic and labour market characteristics, and, crucially, contains no information on educational background.

3.2 How are migrants identified?

3.2.1 Dates of arrival and the transitional arrangements

The period of interest in this study is between May 1st 2004 and April 30th 2011. This covers all migrants arriving after accession, for the entire period of the 'transitional arrangements'. As the arrangements restricted access to government welfare benefits for the first year spent in the UK, they will have affected both the composition of the inflow of A8 migrants, and the labour market behaviour of A8 migrants once in the country. Restricting the analysis to this period ensures that my results describe the characteristics and behaviour of A8 migrants under this specific set of institutional constraints.

The LFS records only the year of arrival of migrants, rather than a specific date or month of arrival. My study identifies all migrants from A8 countries who arrived between 2004 and 2011 as post-accession migrants. As the accession took place on May 1st 2004, this means respondents who arrived between January and April 2004 are misclassified as post-accession migrants. The only feasible alternative to this strategy would be to exclude all A8 workers who arrived in 2004, which would eliminate large numbers of the respondents of interest. For this reason, I proceed with this likely misclassification in mind. Other studies of A8 migrants in the UK have chosen the same strategy (for example, Dustmann *et al.*, 2010). As, at the time of the accession, migrants could not appear in the survey during their first six months in the country, the first post-accession migrants appear in the LFS in November 2004, so I do not include respondents interviewed before this date. The transitional arrangements ended on 1st May 2011, so I also exclude all those interviewed after this date.

3.2.2 Nationality versus country of birth

It is possible to infer migrant status from either the 'nationality' or 'country of birth' variables in the LFS. The literature contains many studies that use 'nationality' to identify migrants, and many that use 'country of birth'. Each of these methods of identification is problematic in its own way. For example, identifying migrants by nationality may cause misclassification because of different naturalisation laws in different countries. Identifying migrants' country of birth may cause misclassification because of nationals being born abroad, particularly in countries with former colonies or military posts abroad, such as the UK (Brücker *et al*., 2002: 72-73).

In this study, I define migrants in terms of their nationality. I see this as a more useful measure of migrant status than country of birth in this case, because nationality is a more fluid concept, which can change over a lifetime as a person moves, or indeed as the borders of a country change. Six of the A8 countries became independent in the early 1990s, which is during the lifetime of many A8 migrants presently in the UK, and therefore a potential source of misclassification in the 'country of birth' variable. Estonia, Latvia, and Lithuania, all part of the former Soviet Union, became independent countries in 1991. Slovenia, part of the former Yugoslavia, became an independent country in 1992, and the Czech Republic and Slovakia, constituting the former Czechoslovakia, separated into independent countries in 1993. A8 migrants born before these dates could potentially report having been born in countries which no longer exist, and, in the cases of the former Soviet Union and the former Yugoslavia at least, be misclassified as non-EU migrants and excluded from the sample.

Table A1 in Appendix A compares the proportion of migrants in each category by the 'nationality' and 'country of birth' definitions'. The only group in the sample for whom the definition makes a substantial difference is the 'EU15' group –17% of whom would have been classified as 'non-EU' migrants had the 'Country of Birth' definition been used. The continents of birth of these particular EU15 nationals are listed in Table A2 (46% were born in African countries, 33% in Asia, and 17% were born in the Americas). This is an interesting finding in itself, but the numbers affected are still relatively small, so I will not pursue the matter here. I will, however, use the 'country of birth' definition of migrant as a robustness check in Appendix D.

3.2.3 Grouping Nationalities

The central comparison in my analysis is between migrants who identify themselves as nationals of one of the A8 countries, who have arrived in the UK since the start of 2004, and migrants who identify themselves as nationals of one of the EU15 countries, who have arrived in the UK over the same time period. Of course, the use of the 'A8' and 'EU15' groups hides much national heterogeneity, but each of these groups is crucially united by a specific set of legal constraints in the UK during the transitional arrangements, so in this case I think the two groupings are useful. Other authors have taken a different approach, such as separating the analysis of Polish migrants (for example, Drinkwater *et al*., 2008).

The countries that are in the European Economic Area (EEA) but not the EU (Iceland, Lichtenstein and Norway) are not included, despite effectively having open borders with the UK – this is because there are slight legal differences in the entitlements of these citizens in the UK. Switzerland is not a member of the EU or the EEA, and is not included for the same reason. Malta and the EU area of Cyprus, which joined the EU at the same time as the A8 countries, are not included, as they did not face equivalent 'transitional arrangements'. Finally, Bulgaria and Romania (the 'A2' countries), which joined the EU in 2007, are not included, as nationals of these countries face their own 'transitional arrangements' until 2014. I will comment on the situation of the A2 nationals briefly in the conclusion.

The nationalities of migrants in the A8 and EU15 groups are shown by country in Tables A3 and A4 in Appendix B. The principle feature of the A8 group is the prevalence of Polish nationals, and, to a lesser extent, Slovakian and Lithuanian nationals: 70% of the A8 sample report Polish nationality, and 88% report either Polish, Slovakian, or Lithuanian nationality. In contrast, the EU15 group features relatively large proportions of several nationalities: respondents from France, Germany, Ireland, Italy and Portugal together make up 71%.

3.3 Sample construction

3.3.1 All occupations

I draw the sample from 27 calendar quarters of the LFS, from the fourth quarter of 2004 to the second quarter of 2011.[14] I exclude respondents outside the ages of 16-64, those who are not employed, those in full-time education, and those reporting nationalities of countries other than UK, A8 or EU15 countries.

Most studies looking at A8 migrants in the UK using the LFS have discarded all but the first wave of the survey. For example, Drinkwater *et al.* (2009: 167-168) favour this approach, as wave one contains earnings information (waves 2-4 do not), and using only the first wave avoids the risk of double counting respondents, avoids mode effects associated with the switch to telephone interviewing after wave one, and avoids attrition bias caused by short-stay migrants leaving the sample prematurely. Response rates are also highest in wave one, at around 70 per cent (ONS, 2011). However, discarding waves 2-5 results in a smaller sample size than would otherwise be available. A common technique to increase sample size is to pool the data from the first wave of the LFS over several years, and this strategy has allowed for larger sample sizes of A8 migrants as the years have passed since accession, but the sample sizes used for analysis in this area of research have so far still been relatively small.

[14] Until the start of 2008, it was not possible for migrants to appear in the sample until they have been in the UK for six months. November 2004 is six months after the EU accession of the A8 countries, and is therefore the first month in which post-accession migrants appear.

I use a novel strategy to exploit the potential of the LFS more fully, allowing a substantial increase in the size of the cross-sectional sample. As noted above in Section 3.1, the LFS follows households rather than respondents, and individual respondents may therefore appear for the first time in any of the five waves. For the same reasons that migrants in general, and A8 migrants in particular, are thought to be under-sampled, they also tend to appear more intermittently across the five waves, and a disproportionate number of migrants are therefore missed when only the first wave is used.

I use one observation per individual respondent, but in order to maximise the number of individuals in the sample, I do not restrict my search for this observation to the first wave. For the first part of this study, in fact, the observation may be drawn from any of the five waves, depending on which waves the individual appears in. The second part of this study requires wage information, and therefore I use only unique observations on individual respondents who provided wage information in waves one or five. The self-employed are automatically excluded in the second part, as wage information is not available on this group.

Table 1 compares the frequency of unique observations with education and occupational information in each nationality group using the first wave only, with that found by augmenting the first wave with individuals observed in the other waves. It also compares the frequency of unique observations that have wage information in the first wave, with that found by adding in observations with wage information only from the fifth wave. 'UK' represents respondents who report UK nationality, 'A8' represents respondents who report being a national of one of the A8 countries, who have come to the UK since 2004, and 'EU15' represents respondents who report being a national of one of the EU15 countries, who have come to the UK in the same time period.

In the construction of the 'Waves 1 and 5' and 'All Waves' samples, I prioritise Wave 1 observations, due to the high response rate and complete set of questions in this Wave. If a respondent is absent from Wave 1, I use the information from Wave 5, as although Wave 5 tends to have the lowest response rate, it still contains earnings information. This constitutes the complete sample for the 'Waves 1 and 5' sample. For the 'All Waves' sample, I also use individuals who do not appear in Waves 1 or 5, adding observations from Waves 2, 3 and 4, in that order, on the grounds that response rates get progressively lower over each wave.

Table 1: Number of individuals from each nationality group in the sample

Nationality	Education and occupational information		Wage information	
	Wave 1	All waves	Wave 1	Waves 1 & 5
UK	258,088	309,240	169,721	190,701
A8	2,987	5,174	2,193	2,940
EU15	967	1,600	654	874
Total	262,042	316,014	172,568	194,515

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education.

Thus, I increase the number of A8 respondents in the sample with education and occupational information by 2,187 (73%) by adding in individuals found in all waves. Similarly, I increase the number of EU15 respondents by 633 (65%) by adding in individuals from all waves, and the number of UK respondents by 51,152 (20%). The number of each group on whom there is wage information also increases. This strategy increases the size of the migrant sample proportionally much more than it increases the size of the native sample, because, as noted above, a higher proportion of migrants miss the first wave and appear in the subsequent waves.

My strategy of constructing a cross-sectional sample using responses from all five waves of the survey could be criticised on the grounds that the method of data collection moves from face-to-face interview to telephone interview for waves two to five, which may alter the way respondents answer some questions. However, there is some evidence that mode effects are less prevalent with objective questions, such as those asked in the LFS (see, for example, Nicolaas *et al.* 2011). My strategy also precludes the use of the sample weights provided with the LFS. However, in my view these disadvantages are outweighed by the benefits of having a substantially larger analytical sample.

3.3.2 'A8 intensive' occupations

New migrant groups are seldom evenly dispersed across different occupations, and in fact are often concentrated in just a few. There are 12 occupations in my sample that have more than one hundred A8 migrants.[15] These are largely unskilled manual or service sector occupations, which might be known in the sociological literature as 'secondary labour market' occupations – that is, they are part of a segmented labour market consisting of occupations that provide low wage, insecure employment with low returns to education, to workers who may face discrimination or other obstacles in the 'primary' labour market. In fact, a high

[15] The 12 occupations are as follows: 'Assemblers and routine operatives', 'Construction trades', 'Elementary cleaning occupations', 'Elementary goods storage occupations', 'Elementary personal service occupations', 'Elementary process plant occupations', 'Food preparation trades', 'Healthcare & related personal service occupations', 'Process operatives', 'Plant and machine operatives', 'Sales assistants and retail cashiers', 'Transport drivers and operatives'.

concentration of A8 workers in secondary labour market occupations would be expected for many of the same reasons that I expect to see a higher risk of over-education in this group (see the discussion of motives, time-horizons, reservation wages, and possible discrimination in Section 2.3). I expect returns to surplus education to be lower in such occupations, since the scope for using surplus education to add productive value is particularly limited. Some of the analysis that follows is restricted to these 'A8 intensive' occupations. Table 2 compares the number of each nationality group in these 12 'A8 intensive' occupations.

Table 2: Number of people in the 'A8 intensive' occupations sample, by nationality group

	Non-A8 intensive	A8 intensive	Total
UK	230,633	78,607	309,240
A8	1,545	3,629	5,174
EU15	1,146	454	1,600
Total	233,324	82,690	316,014

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. Restricted to 12 occupations with at least 100 A8 migrants.

Table 2 shows that 70% of the A8 migrants in the sample are working in these 12 'A8 intensive' occupations, compared to 28% of EU15 migrants, and 25% of UK nationals.

3.4 Measuring qualifications

3.4.1 Using the 'Age completed full-time education' variable

Over the period of interest, the LFS does not capture non-UK qualifications very effectively. In fact, until 2011 all non-UK qualifications were classified as 'Other' in the survey – which presents serious difficulties in assessing relative labour market performance and returns to education for those educated abroad.[16] This is a problem faced by all similar surveys – without the capacity to provide a list of all potential qualifications from every possible country of origin, there is inevitably some inaccuracy in the classification of non-native qualifications.

The tendency in studies of migrants using the LFS has been to use the 'Age completed full-time education' variable as a proxy for the educational achievements of the respondent. Without any further information on why an individual completed education at that particular age, it is very difficult to assess quite what this age might mean for anyone, but it is especially difficult in the case of migrants who have studied in different education systems. The assumption underlying the use of this variable is that a marginal year of education means something roughly equivalent within every country, and indeed at every level of education within

[16] A set of new variables was introduced to the LFS in the first quarter of 2011, in order to capture qualifications gained abroad more effectively.

every country. The measure is therefore supposed to allow the comparison of 'returns' to each year of education for natives and migrants, and for people educated to different levels. However, the diversity of national education systems, even within the European Union, means that that marginal year of education can mean quite different things in different countries. Even the statutory minimum school leaving age varies within the EU. For example, several studies using the 'age left full-time education' variable have noted particularly low returns to education for workers from Poland (e.g. Eade *et al*. 2007, 6; Drinkwater *et al.* 2009: 178). However, the current statutory minimum school leaving age in Poland (18) is two full years higher than that in most EU countries, and three years higher than in some. It is hard to see 'age left full-time education' as an effective measure of educational attainment in these circumstances.

3.4.2 Using ISCED levels

I use a novel strategy to classify educational attainment more accurately for each country, using 'age completed full-time education' to assign respondents to a level in the International Standard Classification of Education (ISCED), a categorisation system that takes account of differences between national education systems (UNESCO, 2006). The different statutory school leaving ages are not as problematic here, as I have classified respondents according to what different school leaving ages should in principle mean in their country of origin. Of course, there is still room for misclassification of respondents' education level, particularly as the survey only records 'age left full-time education' in years, and the fact that some students may be held back to repeat years;[17] this method is by no means equivalent to having direct information on each individual's qualifications. However, this approach does at least exploit the 'age completed full-time education' information fully, by placing it in the context of the relevant education system.[18] Table 3 shows the results of this process.

[17] 'Grade retention' causes measurement problems for the EU15 countries, where in some places (France, Germany, Spain, Portugal for example), 15% or more pupils repeat years at secondary level (EACEA, 2011). I discuss alternative specifications in Appendix F, allowing for large amounts of measurement error in assigned ISCED level for the EU15 group.

[18] I use the tables in the PISA 2009 Technical Report (OECD, 2012a: 364) and the Primary school starting ages from the World Bank's World Development Indicators (World Bank, 2012) to calculate the usual age at which each ISCED level is attained in each country.

Table 3: ISCED level attained, based on age left full-time education, by nationality group (%)

	ISCED levels							
Nationality	**Never had education**	**Primary**	**Lower secondary**	**(Upper) secondary**	**Post-secondary, non-tertiary**	**Vocational tertiary**	**Academic tertiary**	**Total**
UK	0	0	1	59	13	6	21	100
A8	0	1	1	34	16	26	21	100
EU15	0	1	5	24	4	16	51	100
Total	0	0	1	58	13	6	22	100

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. n=316,014.

Comparing these figures to the OECD's estimates for 'Educational attainment: Adult population' (OECD, 2012b) reveals that this method causes me to under-estimate the proportion of people in the UK who have completed tertiary education (the OECD tables suggest that 38% of people in the UK have completed some kind of tertiary education, compared to my estimate of 28%). However, I would expect the OECD figures to be higher, given that they are based on qualifications attained at any time in life, rather than those attained before leaving full-time education.[19] This method also causes me to over-estimate the proportion of people in the UK who have attained upper secondary-level education, which is equivalent to gaining five GCSEs at grade C or above (OECD, 2012b). This may be caused by the end of compulsory schooling in the UK occurring at the same age. The OECD do not publish separate tables for migrants in the UK. Clearly, this method of assigning ISCED levels on the basis of age-left full time education creates some measurement error, but my empirical strategy takes account of this, and is deliberately designed to produce conservative estimates of over-education (see Section 4.1, below).[20]

3.5 Occupation and 'required' education

I classify the occupations of respondents using the 3-digit 'Standard Occupational Classification' (SOC) reported in the LFS. I classify the entire sample using the SOC 2000 definitions, and I use the modal ISCED level of UK nationals within each 3-digit occupation as the 'required' level of education. The modal definition has been used in other migrant over-education papers (Battu and Sloane, 2004: 543; Lindley and Lenton, 2006: 5), with the aim of establishing the standard level of education among workers in each occupation in the absence of migrant effects. This could be problematic if any 3-digit occupation were dominated by migrants, but UK workers make up the great majority of the workforce in every 3-digit occupation in the sample. Using

[19] See Jenkins (2013) for recent evidence on the prevalence of adult learning in the UK

[20] A drawback of assigning ISCED levels based on 'age left full-time education' is that it is sometimes the same for different levels of qualification. I have taken a conservative approach and assigned the lower-level qualification in such cases (see Appendix C).

this approach, the required level of education within every 3-digit occupation is either 'Upper-secondary' or 'Academic tertiary'.

It is possible for respondents to be 'under-educated' using the ISCED definition, but the number of people in the sample who would be classed as such is relatively small, and, for the purposes of this study, the 'under-educated' are considered 'matched'. The relative youth of the migrant groups of interest in this study means few would be classified as under-educated, as such workers tend to be older. For the UK 'under-educated', I assume that additional labour market experience is acting as a substitute for formal education (see Groot, 1999). In these circumstances, it is reasonable to assert that these 'under-educated' workers are 'matched', and the small numbers of such workers in the sample mean that this classification will not affect the results of the study substantively.

Using this method, it is not possible for someone in a 'graduate' occupation to be classed as over-educated. This will cause me to under-estimate the prevalence of over-education, particularly among the UK and EU15 groups. The A8 group will be largely unaffected by this, since so few in the sample work in these occupations. Table 4 shows the number of each nationality group in graduate and non-graduate occupations.

Table 4: Number of people in graduate occupations, by nationality group

	Non-graduate	Graduate	Total
UK	250,321	58,919	309,240
A8	5,018	156	5,174
EU15	1,154	446	1,600
Total	256,493	59,521	316,014
Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education.			

Thus, only 156 (3%) people in the A8 migrants in the sample work in graduate occupations, while 446 (27%) of the EU15 migrants do so, and 58,919 (19%) of the UK nationals. When estimating the wage associations of over-education, I will use only non-graduate occupations, in order to prevent higher average wages among the 'matched' workers in graduate occupations from swamping the wage associations of over-education in the non-graduate occupations. Since wage information is only available for a subset of these workers in non-graduate occupations, the sample size will be further reduced in this part of the analysis.

3.6 Sample definitions

For ease of reference, Table 5 labels the different samples used in the remainder of this study as Samples 'A', 'B', 'C', 'D', and 'E'. Sample B is a sub-set of Sample A, while Samples D and E are the constituent parts of

Sample C. The numbers in the latter three samples are smaller than those with wage information seen in Table 1, as in this case only those in non-graduate occupations are considered.

Table 5: Sample definitions

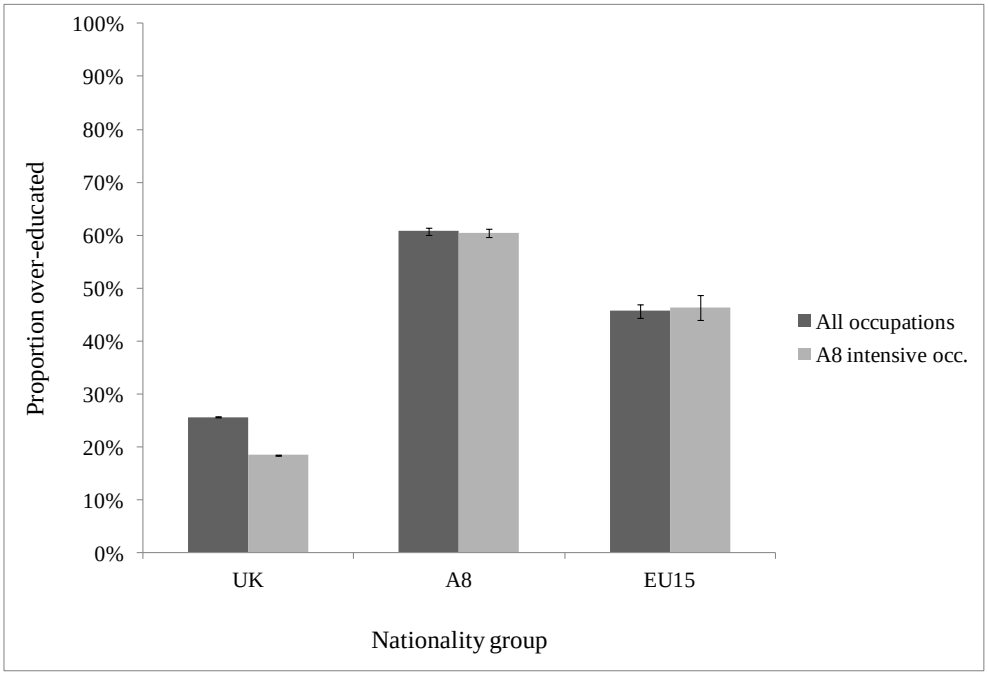
Sample	Description	n
A	All occupations	316,014
B	A8 intensive occupations	82,690
C	Non-graduate occupations, with wage information	157,194
D	Non-A8 intensive, non-graduate occupations, with wage information	107,818
E	A8 intensive, non-graduate occupations, with wage information	49,376

4. Is there any evidence of over-education?

4.1 The prevalence of over-education

Using the definition of 'required' education outlined above, it is possible to compare the prevalence of over-education among the respondents of different nationalities. Figure 2 shows the proportion of each nationality group that is 'Matched' and 'Over-educated', across all occupations, and in the A8 intensive occupations.

Figure 2: Prevalence of over-education, by nationality group



Source: LFS. Notes: Each proportion is a mean value, and bars represent standard error of mean. Men and women, aged 16-64, not in full-time education. Samples A & B in Table 5. n= 316,014 (all occupations), n= 82,690 (A8 intensive occupations).

There is a higher prevalence of over-education among migrants than among natives, with 61% of A8 migrants and 46% of recent EU15 migrants over-educated, compared to 26% of UK nationals. The proportion of UK nationals who are over-educated is lower in the 'A8 intensive' occupations, at 18%. However, restricting the sample to these occupations does not substantially change the prevalence of over-education among either of the migrant groups.

4.2 Can over-education be explained by other observed characteristics?

Differences in the prevalence of over-education between different groups of migrants may be associated with differences in their demographic, human capital, or occupational characteristics. For example, if, in general, younger people face a higher risk of over-education, then a migrant group with a younger age profile will also face a higher risk of over-education, independent of any migrant effects. Alternatively, if part-time workers face a higher risk of over-education, then a migrant group with a higher proportion of part-time workers will also face a higher risk of over-education. I can assess the importance of such factors in explaining the prevalence of over-education in each migrant group by using a probability model, which quantifies the likelihood of respondents being classed as over-educated, conditional on their observed characteristics. However, if I am correct in thinking that the different levels of over-education in A8 and EU15 migrants are driven by unobserved differences between the groups, as well as by possible labour market discrimination, then such a model will not explain all of the difference in the risk of over-education.

Given that I have classified all workers as 'matched' or 'over-educated', which are mutually exclusive categories, I use a probit model here. The model estimates the relative probability of being 'over -educated' (β_1) compared to being 'matched' (β_0). Using the set of parameter estimates, β_1, the relative probability of any individual being over-educated can be calculated, conditional on a vector of observed characteristics, including their nationality group. A positive estimate indicates a higher risk of over-education relative to the reference category, and a negative estimate indicates a lower relative risk of over-education.

I include controls for demographic characteristics, with a gender dummy and a set of four age dummies, for location of workplace, with dummy variables representing the South-east and the 'Regions' (those areas outside London and the South-east), and for job characteristics, with dummy variables for part-time employment, for having 'supervisory responsibilities', and for being in an 'A8 intensive' occupation. For the purposes of comparison with the results, and in order to get a sense of the average characteristics of each nationality group in the sample, Table 6 shows the proportion in each nationality group to which each of these control variables applies, given as percentages.

Table 6: Distribution of control variables, by nationality group (%)

	Nationality group			
Characteristics	**UK**	**A8**	**EU15**	**Total**
Gender				
Female	48	43	46	48
Age				
16-25	13	33	23	14
26-30	10	33	28	11
31-35	11	16	20	11
36-45	25	11	19	24
46-64	41	8	11	40
Place of work				
London	11	14	38	11
South-east	13	11	16	13
Regions	76	75	46	76
Job				
Part-time	23	11	18	22
Supervisor	33	10	26	32
A8 intensive occ.	25	70	28	26

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. Sample A in Table 5. n= 316,014.

Women tend to face a higher risk of over-education than men. It has been suggested that this is because, in a male/female relationship, the higher earnings potential of men gives them greater power in the decision of where to locate (Frank, 1978; Battu, Seaman and Sloane, 1998). This constrains the job-search of women in relationships, and increases the probability that a woman will have to accept a job for which she is over-educated. Discrimination against women may also be a factor (Chevalier, 2003: 517). Table 6 shows that around 43% and 46% of the A8 and recent EU15 migrant groups in the sample is female, while closer to 48% of the UK nationals group is female. This is in line with the evidence from the WRS, cited in Drinkwater *et al*. (2009: 167), which also suggests that around 43% of A8 workers are female. Taking account of gender should therefore increase the relative risk of over-education among both migrant groups.

Younger workers tend to face a higher risk of over-education than older workers because they have had less time to acquire information about the labour market, and may also require more labour market experience to fill any gaps in their skills (Chevalier, 2003: 517). Table 6 shows that 66% percent of A8 migrants in the sample are 30 or younger, compared to 51% of recent EU15 migrants, and 23% of UK nationals. This is in line with the evidence from the WRS, cited in Drinkwater *et al*. (2009: 167). Taking account of age should therefore reduce the relative risk of over-education in both migrant groups, and particularly for A8 migrants.

The theory and evidence on the link between living in an urban area and the risk of over-education is ambiguous (for contrasting empirical results, see Chiswick and Miller, 2009, and Poot and Stillman, 2010), and in this case I argue that the important geographical distinction is between those working in London and the surrounding area, and those working outside it, rather than a general rural/urban divide. Table 6 shows that the employment of recent EU15 migrants is very concentrated around London, with 54% working in either the London or the South East. In contrast, only 25% of A8 migrants work in the same area, a similar proportion to UK natives. I suspect that working in London and the surrounding areas provides strong hedonic benefits for some people, and that these people may accept employment below their level of education in order to capture such benefits. This effect may be particularly strong for migrants, who can potentially benefit from stronger migrant networks and cultural opportunities in the city. Given the high proportion of recent EU15 migrants living in London and the South East, accounting for this geographic distribution is likely to reduce the relative risk of over-education for the recent EU15 migrant group. As A8 migrants are more dispersed, accounting for those working in and around London is not likely to affect their relative risk of over-education substantially.

Part-time employment is generally associated with a higher prevalence of over-education, since part-time work is not possible in all jobs, and those seeking it will therefore have limited opportunities compared to those seeking full-time work. Having supervisory responsibilities may also be associated with a higher prevalence of over-education, if those with more education within a given occupation are more likely to be given such responsibilities. However, including 'part-time employment' and 'supervisory responsibilities' variables in my model potentially introduces some endogeneity problems, since the decision to take part-time work, or the acceptance of supervisory responsibilities may itself reflect the same underlying factors which produce a higher risk of over-education. However, given that different migrant groups can have very different employment profiles, ignoring these factors would introduce a risk of attributing over-education independently associated with these employment characteristics to migrant-specific effects. I will show results of the model without these variables, so that their impact on the estimates of interest is clear.

As noted above, 70% of A8 migrants in the sample work in one of 12 occupations, and Figure 2 showed that the UK workers in these occupations are less likely to be over-educated. Accounting for those working in these occupations is therefore likely to reduce the relative risk of over-education among recent EU15 migrants, but not among A8 migrants, the majority of whom work in these occupations anyway. As this is potentially the most endogenous of the control variables, I will introduce it to the model last.

Table 7 shows the results of the probit model. The control variables are gradually introduced over the columns from left to right, so that the impact of each control on the coefficients of interest is clear. The first column shows the estimates for only the nationality groups, without any control variables, the second column shows the estimates after introducing the gender dummy and the four age dummies, the third column shows the estimates after adding the location dummies to the equation, the fourth column shows the estimates after

adding the controls relating to part-time work and supervisory responsibilities, and the final column shows the estimates after introducing the control for working in an 'A8 intensive' occupation.

The logic behind introducing the control variables in the order is to account for the demographic differences between the groups first, before accounting for the different geographical distribution of the groups, and finally factors associated with the different occupational distribution of the groups. The 'reference person' is a UK national, male, age 16-25, living in London, working full-time, in a non-supervisory role, and in a non-'A8 intensive' occupation.

Table 7: Probit model of over-education, with increasing control variables

	Additional control variables				
Characteristics	**No controls**	**Age/gender**	**Location**	**Job type**	**A8 int. occ.**
Nationality					
(Comp: UK)					
A8	93	70	70	73	87
	(2)	(2)	(2)	(2)	(2)
EU15	55	37	29	30	32
	(3)	(3)	(3)	(3)	(3)
Gender					
Female		12	12	12	10
		(0)	(0)	(1)	(1)
Age					
(Comp: 16-25)					
26-30		1	0	-2	-5
		(1)	(1)	(1)	(1)
31-35		-18	-19	-22	-26
		(1)	(1)	(1)	(1)
36-45		-42	-43	-46	-49
		(1)	(1)	(1)	(1)
46-64		-63	-64	-66	-69
		(1)	(1)	(1)	(1)
Place of work					
(Comp: London)					
South-east			-17	-16	-15
			(1)	(1)	(1)
Regions			-30	-29	-26
			(1)	(1)	(1)
Job					
Part-time				3	7
				(1)	(1)
Supervisor				14	10
				(1)	(1)
A8 intensive occ.					-31
					(1)
Constant term	-65	-35	-11	-14	-5
	(0)	(1)	(1)	(1)	(1)
n	316,014	316,014	315,689	315,689	315,689

Source: LFS. Notes: Standard errors in parentheses. Parameter estimates and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education. Sample A in Table 5.

The estimate on the A8 variable starts out much higher than that on the EU15 variable on the left-hand side of the table, reflecting the higher prevalence of over-education in this group seen in Figure 2. The estimates for the A8 and EU15 variables fall substantially in the second column, after taking account of age and gender. As anticipated, the sign and magnitude of the estimates associated with the age dummies suggest that the risk of over-education reduces with age. Taking account of age therefore explains a part of the over-education seen in both of these migrant groups, which are both youthful relative to the UK comparison group. The higher proportion of A8 nationals in the youngest age categories is reflected in a greater fall in the A8 estimate. The sign of the estimate on the gender dummy is positive, suggesting a higher risk of over-education among women, which is also consistent with my expectations.

The A8 estimate is stable in column 3, after accounting for those that work in London and the South-east, but the EU15 estimate falls again. The sign and magnitude of the estimates associated with the location dummies suggest that, as anticipated, working outside London and the South-east reduces the relative risk of over-education, and that this risk is particularly high in London. As discussed above, a much larger proportion of the EU15 group live in this area, and hence the reduction in the relative risk of over-education in this group after introducing these controls.

Job characteristics are accounted for in column 4, but both nationality estimates are stable. The sign and magnitude of the estimates associated with the occupational dummies suggest that those with 'supervisory' positions are more likely to be over-educated, but part-time work has only a small positive effect. As the 'A8 intensive' occupation dummy is introduced in column 5, the EU15 estimate is stable, but the A8 estimate rises to a level similar to that which it was before introducing controls. This reflects the fact that UK nationals in these occupations face a lower risk of over-education, as is clear from the sign and magnitude of the estimate associated with the A8 intensive dummy. A8 migrants are concentrated in these occupations and yet still face a much higher risk of over-education.

Overall, it appears that much of the risk of over-education among EU15 nationals is associated with their relatively young age profile and the concentration of their employment in London and the South East, but that they still face a relatively high risk of over-education compared to UK nationals. Some of the risk of over-education among A8 nationals is explained by their particularly young age profile, but this group still faces a much higher risk than recent EU15 migrants or UK nationals.

In order to allow a more intuitive interpretation of the magnitude of the effects reported in Table 7, Table 8 shows the impact of a change in each of the variables on the probability of being over-educated in a non-graduate occupation, with all the other variables held at their mean values.

Table 8: Impact of each characteristic on probability of over-education, at mean values, all controls included

Characteristics	
Nationality	
(Comp: UK)	
A8	28
	(1)
EU15	10
	(1)
Gender	
Female	3
	(0)
Age	
(Comp: 16-25)	
26-30	-2
	(0)
31-35	-8
	(0)
36-45	-16
	(0)
46-64	-22
	(0)
Place of work	
(Comp: London)	
South-east	-5
	(0)
Regions	-8
	(0)
Job	
Part-time	2
	(0)
Supervisor	3
	(0)
A8 intensive occ.	-10
	(0)
n	315,689

Source: LFS. Notes: Standard errors in parentheses. Impacts and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education. Sample A in Table 5, with a slightly smaller sample size, as all controls are included here.

For A8 migrants, the impact of their nationality group is to increase the probability of being over-educated by 28 percentage points relative to UK nationals, holding their geographical, occupational and demographic characteristics constant at mean values. The magnitude of this effect is very similar to that implied above in Figure 2, and this reflects that fact that taking account of observed characteristics does not much reduce the relative risk of over-education for A8 migrants. For recent EU15 migrants, the effect of their nationality on the relative probability of being over-educated is lower, at 10 percentage points. The magnitude of this effect is much smaller than that implied in Figure 2, as a fairly large portion of the elevated risk of over-education in this group can be explained by their relative youth, and especially by their geographical concentration in London and the South East.

These results are consistent with my assertion that different levels of over-education are driven by unobserved differences between the groups, arising from distinct self-selection processes associated with the institutional context of the EU accession, as well as by possible differential labour market discrimination. However, the high level of over-education among migrants is only a concern in itself if it is associated with negative outcomes. One potential negative outcome which is relatively straightforward to observe in survey data is reduced earnings potential. The next section investigates whether over-education among A8 migrants is associated with any wage differences.

5. Is over-education associated with any wage effects?

5.1 Is there any evidence of wage differences?

If over-educated workers within an occupation are more productive than matched workers, then they may be paid higher wages. Indeed, previous empirical work suggests that the over-educated are usually paid more than those who are matched to their jobs within the same occupation (Piracha and Vadean, 2012: 21). However, at the same time, recent migrants tend to be paid less than native workers (Chiswick, 1978). The size of these opposing effects varies in different countries and with different migrant groups. In this section I will investigate the size of these effects for A8 and recent EU15 migrants in the UK.

The association of over-education and wages is assessed here based on the income information available for a sub-section of the sample drawn from the LFS. The widely used 'HOURPAY' variable (average gross hourly pay) is used. This is a derived variable, based on reported gross weekly pay, basic usual hours, and usual hours of paid over-time per week. As, by the definitions I use here, no worker in a graduate occupation can be over-educated, I also exclude those graduate occupations from the wage analysis. As noted above, this only excludes 3% of the A8 migrants in the sample, while it excludes 27% of the EU15 migrants and 19% of the UK nationals. The self-employed are not asked to provide wage information, so they are automatically excluded from this part of the analysis as well. These exclusions should be borne in mind when considering the results.

Figure 3 shows kernel density estimates of the log wage distribution for each nationality group, for all non-graduate occupations, and for A8 intensive occupations.[21] Across all non-graduate occupations, the distribution of wages among recent EU15 migrants is distinct from that of UK workers, peaking at a lower point and tailing off more rapidly, until the very top of the distribution where the highest paid EU15 workers are paid more than UK workers. However, the A8 wage distribution in non-graduate occupations is clearly very different to that of both EU15 workers and UK nationals. The peak of the distribution is in approximately the same area of that of the EU15 migrants, but there is a much higher density of A8 migrants at this point, and the distribution tails off very rapidly after this peak.

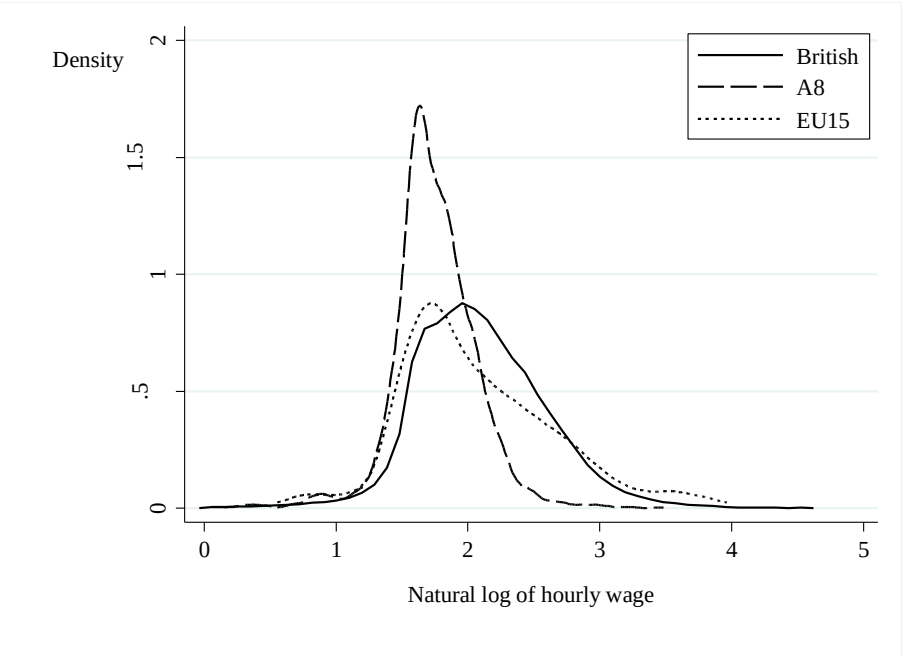
In the 'A8 intensive' occupations, by contrast, the wage distributions of the two migrant groups are almost indistinguishable. The peak of the wage distribution for UK nationals is in

[21] Wages are adjusted for price inflation monthly using the Retail Prices Index (ONS, 2012), with November 2004 as the reference month.

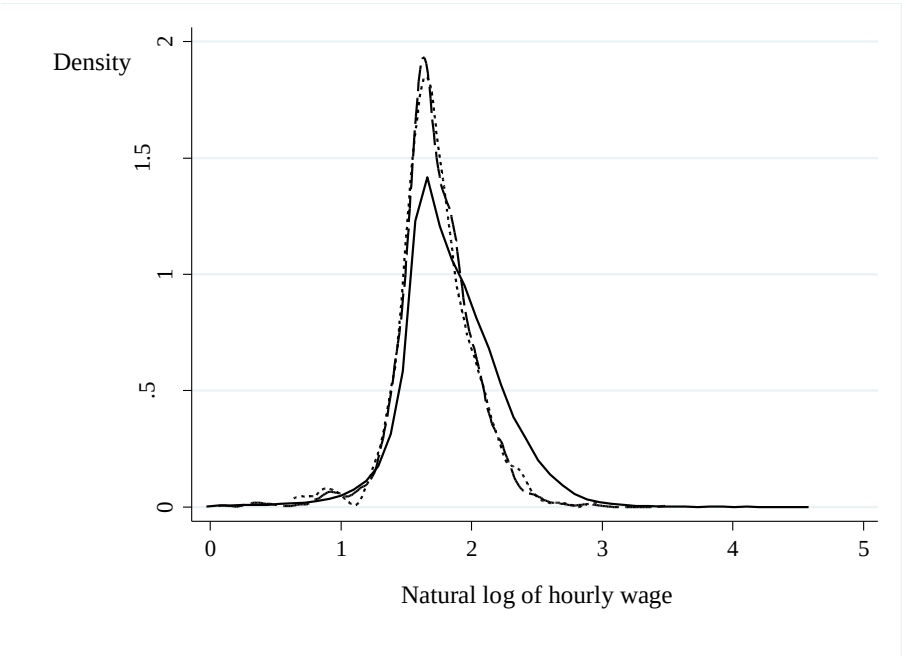
a similar place to that for the two migrant groups, but at a lower density. The much narrower wage distribution in the second part of Figure 3 is characteristic of the kind of secondary labour market occupations in which A8 workers are concentrated. Whether these wage similarities remain after controlling for observed characteristics, and whether there is any association with over-education, is discussed in the next section.

Figure 3: Kernel density estimates of log wage distribution in non-graduate occupations, by nationality group

a) All occupations



b) A8-intensive occupations



Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education, in non-graduate occupations. Sample C and E in Table 5. Sample sizes are slightly lower here as I have excluded those reporting hourly wages of less than £1. n=156,822 (all occupations), n= 49,243 (A8 intensive occupations).

5.2 Can wage differences be explained by other observed characteristics?

In order to establish whether some part of the large discrepancy in non-graduate wages between the nationality groups is predicted by the differences in the prevalence of over-education, I estimate a novel variant of the 'ORU' ('Over-, Required-, and Under-education') wage equations, first used by Duncan and Hoffman (1982). This version of the equation allows for variety in the educational systems of different countries, and also allows for over-education to interact with different nationality groups in distinct ways.

The equation used here takes the form:

$$\ln(w_i) = \alpha + \beta_1 A8_i + \beta_2 EU15_i + \beta_3 OVER_i + \beta_4 A8_i \cdot OVER_i + \beta_5 EU14_i \cdot OVER_i + \gamma X_i + u_i \quad (1)$$

where w_i represents the hourly wage of individual i, *A8* and *EU15* are dummy variables representing each nationality group, *OVER* is a dummy variable representing over-education, and *A8·OVER* and *EU15·OVER* are interaction terms for the joint effect of nationality and over-education. *X* represents a vector of control variables, including year and quarter dummies, which are intended to take account of other factors associated with wage outcomes. As in Figure 3 above, wages are adjusted for price inflation monthly using the Retail Prices Index (ONS, 2012b), with November 2004 as the reference month.

Each of β_1 and β_2 therefore represent the wage effect for matched workers of being an A8 or recent EU15 migrant in a non-graduate occupation, as opposed to a UK national, while β_3 represents the hourly wage returns to over-education for UK nationals. The interaction terms represent any additional effects that arise from being in a particular nationality group and being over-educated – for example, if being both an A8 and over-educated (β_4) has an additional effect beyond the sum of the effect of being A8 and the effect of being over-educated ($\beta_1+\beta_3$). These interaction terms are a useful addition to the standard ORU wage equations, as they separate out any additional differences in wages associated with over-education specific to each nationality group.

This specification also differs from that used in most of the migrant over-education literature (see Piracha and Vadean, 2012: 13-14) in that it estimates a rate of return to the *state* of over-

education, rather than a rate of return to *years* of over-education.[22] The more widely used specification constrains the wage effect of a marginal year of education to be the same whichever country that education comes from, and at whichever level of education that marginal year occurs. As discussed above, this approach does not fit comfortably with the heterogeneity of European education systems. In my approach, the state of being over-educated already takes into account different European education systems, via the ISCED classification system discussed in Section 3.4.2 above, and therefore avoids the problematic notion of a continuous rate of return to over-education across individuals educated in different countries, and across different levels of education within each country.

Given the strong negative association between being in an 'A8 intensive' industry and the likelihood of being over-educated found in the probit model estimated above, these industries are separated in the analysis. Table 9 compares the coefficients resulting from estimating Equation 1 for: (i) All non-graduate occupations, (ii) non-A8 intensive occupations, and (iii) A8-intensive occupations. The reference respondent is a matched UK national, and only controls for year and quarter are included at this stage.

[22] The use of dummies in the estimation of wage effects in the over-education literature comes from Verdugo and Verdugo (1989). Battu and Sloane (2004) take a similar approach.

Table 9: Log wage equations: Nationality and over-education, non-graduate occupations only

Coefficient	(i) All non-graduate occupations	(ii) Non-A8 intensive occupations	(iii) A8 intensive occupations
Over-educated	22	19	1
	(0)	(0)	(0)
A8	-33	-44	-15
	(2)	(3)	(1)
EU15	-20	-15	-12
	(3)	(5)	(3)
Over-educated*A8	-13	0	3
	(2)	(4)	(2)
Over-educated* EU15	16	22	1
	(4)	(6)	(5)
n	157,194	107,818	49,376

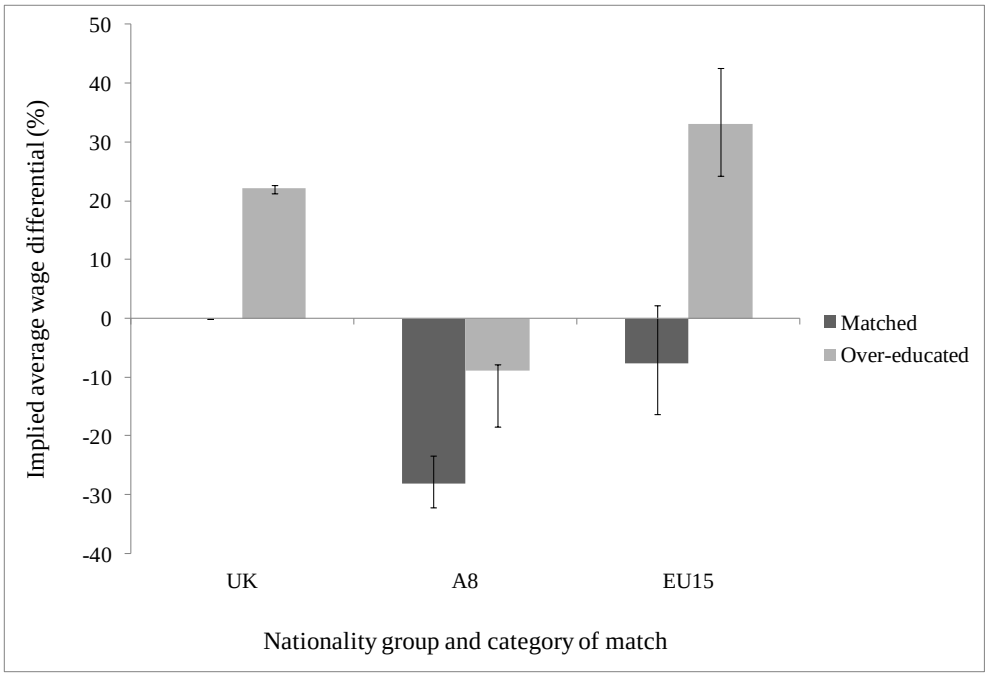
Source: LFS. Notes: Standard errors in parentheses. Coefficients and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education, in non-graduate occupations. Sample C, D, and E in Table 5.

The coefficients for 'Non-A8 intensive occupations' are in general much closer in magnitude to those for 'All occupations' then are those for the 'A8 intensive' occupations. I therefore separate the 'A8 intensive' occupations from the 'Non-A8 intensive' occupations' in the wage analysis.

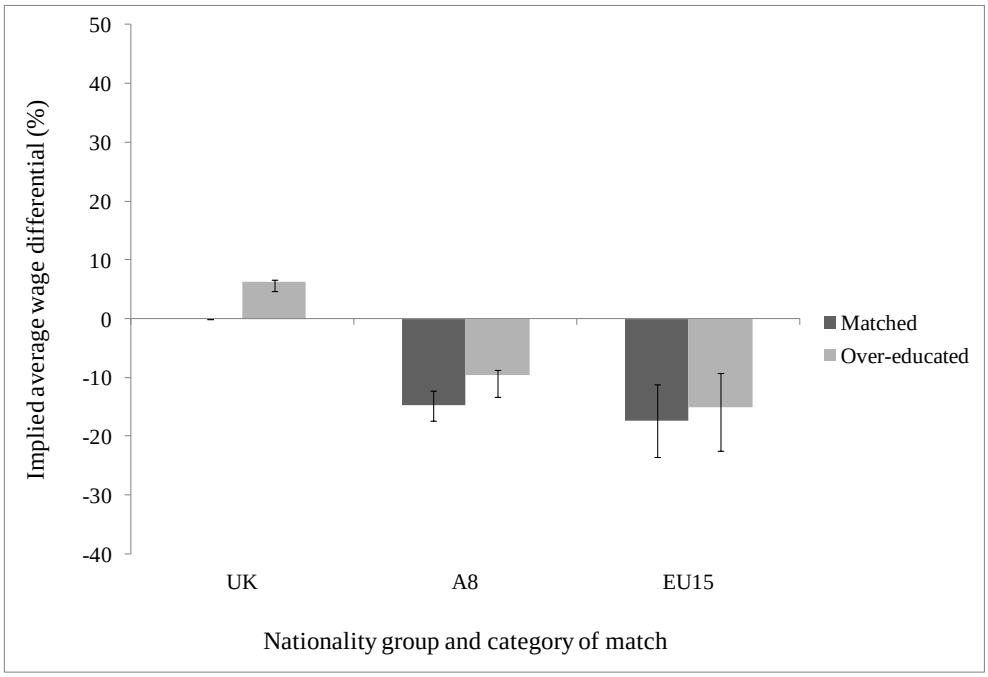
Figure 4 presents the implied average wage differentials from estimating Equation 1 with the full set of controls for non-A8 intensive and A8 intensive occupations. The 'reference person' is a matched UK national who is male, age 16-25, living in London, working full-time, and in a non-supervisory role. The 'matched' bars for A8 and EU15 nationals can be thought of as representing the 'standard' wage penalties for each migrant group, while the 'over-educated' bars can be considered the returns to surplus education for each group.

Figure 4: Implied average wage differentials, compared to matched UK nationals

a) Non-A8 intensive occupations



b) A8 intensive occupations



Source: LFS. Notes: Each proportion is a mean value, and bars represent 95% confidence intervals. Employed men and women, aged 16-64, not in full-time education, in non-graduate occupations. Samples D & E in Table 5. Sample sizes are slightly lower as the equations which produced these charts include the full-set of control variables. n=107,712 (Non-A8 intensive occupations), n= 49,354 (A8 intensive occupations).

Very different stories emerge about the returns to over-education in non-graduate occupations for workers in the two sets of occupations above: Figure 4(a) shows large wage penalties for both migrant groups, partly or fully compensated for by strong positive returns to over-education, while Figure 4(b) shows occupations where wage penalties for A8 migrants are less severe, but where there are almost no compensating returns to over-education available.

Tables 10 and 11 show the gradual introduction of control variables into Equation 1, for non-A8 intensive and A8 intensive occupations respectively (see Appendix E for the equivalent table for all occupations). The control variables include a full set of year and quarter dummies, and, as in the probit model in Section 4.2 above, controls for age and gender, location, and job characteristics. I also include two terms which interact the gender variable with the A8 and EU15 nationality indicators, in order to capture any additional association between gender and wages that is specific to each migrant group. I do this because gender potentially has a smaller effect on wages in the A8 intensive occupations, where the wage distribution is narrower. As in the probit analysis above, I introduce the control variables incrementally, so that it is possible to get a sense of which characteristics are contributing most to each coefficient of interest. The first column only controls for year and quarter, the second column controls for demographic characteristics, the third column controls for location, and the fourth column introduces controls related to occupation.

Table 10: Log wage equations: Non-A8 intensive occupations

	Additional control variables			
Characteristics	**Year/quarter**	**Age/gender**	**Location**	**Job type**
Mismatch				
Over-educated	19	24	22	20
	(0)	(0)	(0)	(0)
Nationality (Comp: UK)				
A8	-44	-40	-40	-33
	(3)	(3)	(3)	(3)
EU15	-15	-6	-12	-8
	(5)	(5)	(5)	(5)
Over*A8	0	-4	-3	-3
	(4)	(4)	(4)	(4)
Over*EU15	22	18	17	17
	(6)	(6)	(5)	(5)
Gender				
Female		-21	-21	-13
		(0)	(0)	(0)
Female*A8		18	14	7
		(3)	(3)	(3)
Female*EU15		0	-1	-5
		(5)	(5)	(4)
Age (Comp: 16-25)				
26-30		29	29	25
		(1)	(1)	(1)
31-35		41	40	37
		(1)	(1)	(1)
36-45		47	47	42
		(0)	(0)	(0)
46-64		45	44	41
		(0)	(0)	(0)
Place of work (Comp: London)				
South-east			-22	-20
			(1)	(1)
Regions			-28	-27
			(0)	(0)
Job				
Part-time				-15
				(0)
Supervisor				23
				(0)
Constant term	213	184	209	201
	(1)	(1)	(1)	(1)
n	107,818	107,818	107,712	107,712

Source: LFS. Notes: Source: LFS. Notes: Standard errors in parentheses. Coefficients and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education, in non-graduate, non-A8 intensive occupations. Sample D in Table 5.

Table 11: Log wage equations: A8-intensive occupations

	Additional control variables			
Characteristics	**Year/quarter**	**Age/gender**	**Location**	**Job type**
Mismatch				
Over-educated	1	7	6	6
	(0)	(0)	(0)	(0)
Nationality (Comp: UK)				
A8	-15	-17	-17	-16
	(1)	(2)	(2)	(2)
EU15	-12	-20	-23	-19
	(3)	(4)	(4)	(4)
Over*A8	3	-2	-1	-1
	(2)	(2)	(2)	(2)
Over*EU15	1	-2	-2	-4
	(5)	(5)	(5)	(5)
Gender				
Female		-22	-22	-17
		(0)	(0)	(0)
Female*A8		15	15	11
		(2)	(2)	(2)
Female*EU15		20	21	17
		(5)	(5)	(5)
Age (Comp: 16-25)				
26-30		17	17	16
		(1)	(1)	(1)
31-35		21	21	20
		(1)	(1)	(1)
36-45		25	25	23
		(1)	(1)	(1)
46-64		23	23	22
		(0)	(0)	(0)
Place of work (Comp: London)				
South-east			-9	-9
			(1)	(1)
Regions			-12	-12
			(1)	(1)
Job				
Part-time				-11
				(0)
Supervisor				11
				(0)
Constant term	185	177	188	187
	(1)	(1)	(1)	(1)
n	49,376	49,376	49,354	49,354

Source: LFS. Notes: Source: LFS. Notes: Standard errors in parentheses. Coefficients and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education, in non-graduate, A8 intensive occupations. Sample E in Table 5.

The coefficients on the over-education and nationality variables in Tables 10 and 11 can be interpreted relative to the reference category, but to calculate the effect of nationality and over-education together, the coefficients on the nationality variables and over-education variables must be summed with the coefficients on the interaction term. Also, as the dependent variable is the natural logarithm of the hourly wage, the antilog must be taken to get a precise percentage interpretation.[23]

Before accounting for any differences in characteristics, in the non-A8 intensive occupations analysed in Table 10, over-educated workers overall earn an average wage premium of 21%, while the average wage penalty faced by A8 migrants in these occupations is 36%. There appears to be no additional effect captured by the A8*over-educated variable, so, ignoring the small and poorly determined coefficient on the interaction term, an over-educated A8 migrant in one of these occupations will earn 15% less than the reference individual, a matched UK national. As different characteristics are accounted for in the table, the positive return to over-education remains at around 20%, becoming 22% after introducing all the controls, while the wage penalty faced by A8 migrants falls to 28% on average, with most of the fall being accounted for by demographic and occupational characteristics. Thus we see that the positive returns to over-education are able to compensate in part for the wage penalty suffered by A8 migrants in these occupations, but that wage penalty is of such a large magnitude that, even after controlling for the characteristics here, such over-educated migrants in these occupations on average still earn 6% less than matched UK nationals.

The positive returns to over-educated migrants from EU15 countries in these occupations are able to fully compensate for the wage penalty that they face. Before introducing controls, recent EU15 migrants in these occupations earn 14% less than UK workers on average, but those recent EU15 migrants who are over-educated earn the over-education premium of 21%, plus an EU15-specific wage premium for over-educated workers of 25%. Overall, over-educated recent EU15 migrants in these occupations earn an average wage premium of 32%. After accounting for the characteristics above, over-educated EU15 workers earn a wage premium of 33% relative to matched UK nationals, with the increase being largely accounted for by their age and gender profile.

In the non-A8 intensive occupations analysed in Table 10, matched A8 migrants face an especially large wage penalty compared to matched UK nationals. However, the strong

[23] The equation to apply to each coefficient to get a percentage interpretation is (100*[exp(β)-1]), where β is the coefficient of interest. Where multiple coefficients apply they must be summed before the antilog is taken (for example, the percentage wage effect for an over-educated EU15 migrant would be (100*[exp(β_2+β_3+β_5)-1]).

positive returns to over-education can at least partly compensate for the wage penalties that migrants face in the UK. For over-educated A8 migrants though, these strong positive returns to over-education are still not big enough to bring their average wages up to the level of matched UK nationals.[24] Furthermore, it is the results of the wage equations represented in Table 11 that reflect the experience of most A8 migrants in the sample, 70% of whom work in these 'A8 intensive' occupations.

In contrast to the occupations analysed in Table 10, there appear to be very low positive returns to over-education in the 'A8 intensive' occupations analysed in Table 11, and the positive return of 6% only appears after taking account of all the observed characteristics above. However, the wage penalty faced by A8 workers is much smaller here, at 15%, and in fact recent EU15 migrants pay a similar penalty of 17% on average, after taking account of observed characteristics. The control variables still generally have a well-determined influence on wages, but the size of the effect for each variable is much smaller. Age is really the only factor in these industries that seems to have a large and well-determined effect.

The wage effects estimated in Table 11 fit well with the idea that these 'A8 intensive' occupations are part of a 'secondary labour market'. Almost no returns to over-education are available to workers of any nationality in these occupations, but wage penalties faced by matched and over-educated migrant workers are much smaller at the same time. These wage penalties may be smaller simply because most employers are bound by the National Minimum Wage in the UK, so there is a 'lower bound' beyond which wages cannot fall, or because even the highest wages paid in these occupations do not rise far above the average wages earned by migrants.

[24] The coefficients on the female*A8 interaction terms are positive and well determined in both Tables 10 and 11, which suggests women migrants earning similar wages to their male counterparts. This is an interesting finding and further investigation in this area may be fruitful.

6. Conclusions

In this paper I have presented the first quantitative evidence on the prevalence and wage associations of over-education among A8 migrants in the UK. I have estimated that 61% of A8 migrants in the UK are over-educated for their jobs, and that very little of this over-education is explained by their observed characteristics. In comparison, I have estimated that 46% of recent EU15 migrants in the UK are over-educated, and that most of this over-education is explained by their age profile and geographical distribution. I have argued that these results are driven by unobserved differences between the groups, arising from distinct self-selection processes associated with the institutional context of the EU accession, and that differential labour market discrimination may also play a part.

In non-graduate occupations, I also examined the association between over-education and wages, and found that overall the over-educated tend to earn more than their peers within each nationality group, and that in some occupations, positive rewards to over-education can partly or wholly compensate the over-educated for average migrant wage penalties. However, the majority of A8 migrants work in occupations where penalties for migrant workers are less severe, and where over-education is barely rewarded with higher wages at all.

With the end of the transitional arrangements in May 2011, A8 migrants in the UK were granted the same access to welfare benefits as other EU migrants, and the final legal restrictions on such migrants working in other EU countries were removed. Over the transitional period, there was also some degree of convergence between real wages available in the A8 countries, and those available in the richer EU15 and the UK. With welfare benefits easing job-search, with a broader selection of alternative destinations from which to choose, and with better prospects at home, the relative prevalence of over-education among A8 migrants in the UK is likely to fall.

In anticipation of the 'A2' countries (Romania and Bulgaria) being granted full access to the UK labour market on January 1st 2014, the Prime Minister of the UK has announced plans to restrict access to various benefits and public services for all new migrants.[25] I have not directly sought to identify the effect of excluding new migrants from government welfare benefits here, nor have I suggested that domestic government policy is the major factor driving over-education among A8 migrants. However, as I have noted above, intuitively, while such benefit exclusions increase employment rates in the migrant stock, they also

[25] See Cameron (2013).

increase the pressure on skilled migrant workers from lower-income countries to take humble jobs in the UK. If policy-makers wish to discourage future migrants from taking up such employment, exclusions from welfare benefits, however politically expedient, are almost certainly counter-productive.

Bibliography

Allen, J. and Van der Velden, R. (2001). 'Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on the job search'. *Oxford Economic Papers*, 53 (3), 434-452.

Anderson, B., Clark, N. and Parutis, V. (2007). *New EU Members?: Migrant Workers' Challenges and Opportunities to UK Trades Unions: Polish and Lithuanian Case Study*. London, UK: Trades Union Congress.

Anderson, B., Ruhs, M., Rogaly, B. and Spencer, S. (2006). *Fair Enough?: Central and East European Migrants in Low-wage Employment in the UK*. York, UK: Joseph Rowntree Foundation.

Battu, H., Seaman, P. & Sloane, P. (1998). 'Are Married Women Spatially Constrained? A test of gender differentials in labour market outcomes'. University of Aberdeen, Department of Economics Working Papers, 98-07.

Battu, H., Seaman, P. and Zenou, Y. (2011). 'Job contact networks and the ethnic minorities'. *Labour Economics*, 18 (1), 48-56.

Battu, H. and Sloane, P. (2004). 'Over-education and ethnic minorities in Britain'. *The Manchester School*, 72 (4), 535-559.

Bell, B., Machin, S. and Fasani, F. (2010). 'Crime and immigration: Evidence from large immigrant waves'. CEP Discussion Paper, 984.

Blackaby, D., Leslie, D., Murphy, P. and O'Leary, N. (2005). 'Born in Britain: How are native ethnic minorities faring in the British labour market?'. *Economics Letters*, 88 (3), 370-375.

Blanchflower, D. G. and Shadforth, C. (2009). 'Fear, Unemployment and Migration'. *The Economic Journal*, 119 (535), F136-F182.

Brücker, H., Epstein, G. S., McCormick, B., Saint-Paul, G., Venturini, A. and Zimmermann, K. F. (2002). *Managing migration in the European welfare state*. Oxford, UK: Oxford University Press.

Cameron, D. (2012). 'Prime Minister's speech on immigration'. [Online]. Available at: https://www.gov.uk/government/speeches/david-camerons-immigration-speech. [Last accessed 29.04.2013].

Cameron, D. (2013). 'David Cameron's immigration speech'. [Online]. Available at: https://www.gov.uk/government/speeches/prime-ministers-speech-on-immigration. [Last accessed 29.04.2013].

Chevalier, A. (2003). 'Measuring over-education'. *Economica*, 70 (279), 509-531.

Chiswick, B. (1978). 'The effect of Americanization on the earnings of foreign-born men'. *The Journal of Political Economy*, 86 (5), 897-921.

Chiswick, B. (2001). 'Are immigrants favorably self-selected? An economic analysis'. IZA Discussion

Paper, 131.

Chiswick, B. and Miller, P. (2009). 'The international transferability of immigrants' human capital'. *Economics of Education Review,* 28 (2), 162-169.

Clark, K. and Drinkwater, S. (2008). 'The labour-market performance of recent migrants'. *Oxford Review of Economic Policy,* 24 (3), 495-516.

Clegg, N. (2013). 'Nick Clegg speech on immigration'. [Online]. Available at: http://www.libdems.org.uk/latest_news_detail.aspx?pPK=e3347217-1fa7-4f09-9a5c-bfb4a716b9df&title=Nick_Clegg_speech_on_immigration. [Last accessed 29.04.2013].

Cook, J., Dwyer, P. and Waite, L. (2011a). 'The experiences of accession 8 migrants in England: motivations, work and agency'. *International Migration,* 49 (2), 54-79.

Cook, J., Dwyer, P. and Waite, L. (2011b). ''Good relations' among neighbours and workmates? The everyday encounters of Accession 8 migrants and established communities in Urban England'. *Population, Space and Place,* 17 (6), 727-741.

Coombes, M., Champion, T. and Raybould, S. (2007). 'Did the early A8 in-migrants to England go to areas of labour shortage?'. *Local Economy,* 22 (4), 335-348.

Drinkwater, S., Eade, J. and Garapich, M. (2009). 'Poles apart? EU enlargement and the labour market outcomes of immigrants in the United Kingdom'. *International Migration,* 47 (1), 161-190.

Duncan, G. and Hoffman, S. (1982). 'The incidence and wage effects of overeducation'. *Economics of Education Review,* 1 (1), 75-86.

Dustmann, C., and Fabbri, F. (2003). 'Language proficiency and labour market performance of immigrants in the UK.' *The Economic Journal,* 113 (489), 695-717.

Dustmann, C., Fabbri, F. and Preston, I. (2005). 'The Impact of Immigration on the British Labour Market'. *The Economic Journal,* 115 (507), F324-F341.

Dustmann, C., Frattini, T. and Halls, C. (2010). 'Assessing the Fiscal Costs and Benefits of A8 Migration to the UK'. *Fiscal Studies,* 31 (1), 1-41.

Dustmann, C., Tommaso F., and Preston, I. (2013). 'The effect of immigration along the distribution of wages'. *The Review of Economic Studies,* 80 (1), 145-173.

Dustmann, C., and Weiss, Y. (2007). 'Return migration: theory and empirical evidence from the UK'. *British Journal of Industrial Relations,* 45 (2), 236-256.

Eade, J., Drinkwater, S. and Garapich, M. (2007). *Class and Ethnicity: Polish Migrants in London: Full Research Report.* Swindon: ESRC.

Engbersen, G., Snel, E. and de Boom, J. (2010). 'A 'van full of poles': liquid migration from Central and Eastern Europe'. In R. Black, G. Engbersen, M. Okólski and C. Pantīru (Eds.), *EU Enlargement and labour migration from Central and Eastern Europe* (115-40). Amsterdam, Netherlands: Amsterdam University Press.

Education, Audiovisual and Culture Executive Agency (EACEA) (2011). *Grade Retention during Compulsory Education in Europe: Regulations and Statistics.* Brussels: EACEA.

Eurostat (2013). *Gross domestic product at market prices*. [Online]. Available at: ec.europa.eu/eurostat/product?code=tec00001. [Last accessed 28.04.2013].

Fomina, J. and Frelak, J. (2008). *Next stopski London: Public perceptions of labour migration within the EU. The case of Polish labour migrants in the British press*. Warsaw: Institute of Public Affairs.

Frank, R. (1978). 'Why women earn less: the theory and estimation of differential overqualification'. *The American Economic Review,* 68 (3), 360-373.

Gilpin, N. (2006). 'The impact of free movement of workers from Central and Eastern Europe on the UK labour market'. Working Paper 29, Department for Work and Pensions.

Green, C., Kler, P. and Leeves, G. (2007). 'Immigrant overeducation: Evidence from recent arrivals to Australia'. *Economics of Education Review,* 26 (4), 420-432.

Groot, W. and van den Brink, H. (1999). 'Job satisfaction of older workers'. *International Journal of Manpower,* 20 (6), 343-360.

Jayaweera, H., Anderson, B. and Phil, D. (2008). *Migrant workers and vulnerable employment: a review of existing data*. Oxford: Centre on Migration, Policy and Society (COMPAS).

Jenkins, A. (2013). 'Learning and the lifecourse: The acquisition of qualifications in adulthood'. CLS Cohort Studies Working Paper, 2013/4.

Lemos, S. and Portes, J. (2008). 'The impact of migration from the new European Union Member States on native workers'. *Department for Work and Pensions Working Papers*, 52.

Lindley, J. (2009). 'The over-education of UK immigrants and minority ethnic groups: Evidence from the Labour Force Survey'. *Economics of Education Review,* 28 (1), 80-89.

Lindley, J. and Lenton, P. (2006). 'The over-education of UK immigrants: Evidence from the Labour Force Survey'. Sheffield Economic Research Paper Series, 2006001.

Migration Advisory Committee (2012). *Analysis of the Impacts of Migration*. London, UK: Migration Advisory Committee.

Miliband, E. (2013). 'Ed Miliband: Immigration Must Work For All'. [Online video]. Available at http://www.youtube.com/watch?v=vDELSPXMwIY. Last accessed 29.04.2013.

Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A. and Lynn, P. (2011). 'Is it a good idea to optimise question format for mode of data collection? Results from a mixed modes experiment'. ISER Working Paper Series, 2011-31.

Office for National Statistics (2004-2011). 'Quarterly Labour Force Survey'. [Computer files]. Colchester, Essex, UK: UK Data Service.

Office for National Statistics (2011). *Labour Force Survey User Guide*. (Vol. 1: Background and Methodology). London, UK: Office for National Statistics.

Office for National Statistics (2012a). *National Population Projections, 2010-based reference volume*. London, UK: Office for National Statistics.

Office for National Statistics (2012b). *Consumer Price Indices*. London, UK: Office for National

Statistics.

Office for National Statistics (2013a). *Provisional Long Term International Migration estimates*. London, UK: Office for National Statistics.

Office for National Statistics (2013b). *Labour Market Statistics, April 2013*. London, UK: Office for National Statistics.

Organisation for Economic Co-operation and Development (2012a). *PISA 2009 Technical Report*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (2012b), *Education at a Glance 2012: OECD Indicators*. Paris, France: OECD Publishing.

Parutis, V. (2011). '"Economic Migrants" or "Middling Transnationals"? East European Migrants' Experiences of Work in the UK'. *International Migration*, 1-31.

Pemberton, S. (2009). 'Economic migration from the EU 'A8' accession countries and the impact on low-demand housing areas: opportunity or threat for Housing Market Renewal Pathfinder programmes in England?'. *Urban Studies*, 46 (7), 1363-1384.

Piracha, M. and Vadean, F. (2012). 'Migrant educational mismatch and the labour market'. IZA Discussion papers, 6414.

Poot, J. and Stillman, S. (2010). 'The importance of heterogeneity when examining immigrant education-occupation mismatch: evidence from New Zealand'. CReAM Discussion Paper Series, 1023.

Portes, J. and French, S. (2005). 'The impact of free movement of workers from central and eastern Europe on the UK labour market: early evidence'. Department for Work and Pensions Working Papers, 18.

Robinson, D. (2007). 'European Union Accession State Migrants in Social Housing in England'. *People, Place & Policy Online*, 1 (3), 98-111.

Salt, J. and Rees, P. (2006). *Globalisation, population mobility and impact of migration on population*. Swindon, UK: Economic and Social Research Council.

Shields, M. and Wheatley-Price, S. (1998). 'The earnings of male immigrants in England: evidence from the quarterly LFS'. *Applied economics*, 30 (9), 1157-1168.

Spencer, S., Ruhs, M., Anderson, B. and Rogaly, B. (2007). *`Migrants' lives beyond the workplace: the experiences of Central and East Europeans in the UK'*. York, UK: Joseph Rowntree Foundation.

Sumption, M. (2009). 'Social networks and Polish immigration to the UK'. IPPR, Economics of Migration Working Paper, 5.

Sumption, M. and Somerville, W. (2010). *The UK's new Europeans*. London, UK: Equality and Human Rights Commission.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2006). '*International Standard Classification of Education, ISCED 1997*'. Paris, France: UNESCO.

Verdugo, R. and Verdugo, N. (1989). 'The impact of surplus schooling on earnings'. *Journal of Human Resources,* 24 (4), 629-643.

Weishaar, H. (2008). 'Consequences of international migration: A qualitative study on stress among Polish migrant workers in Scotland'. *Public health,* 122 (11), 1250-1256.

World Bank (2013). *Primary school starting age (years)*. [Online]. Available at: http://data.worldbank.org/indicator/SE.PRM.AGES. [Last accessed 29.04.2013].

Appendix A: Comparing the 'Nationality' and 'Country of Birth' definitions of migrant.

The main body of this paper uses a 'Nationality' definition of migrant, based on each respondent's reported nationality. Table A1 compares the proportion of each nationality group that would be classified differently using a 'Country of birth' definition of migrant, based on each respondent's reported country of birth. Table A2 examines the continents of origin for those EU15 nationals in the UK who were born outside the EU.

Table A1: Percentage of each 'nationality group' that would be classified differently using a 'Country of Birth' definition of migrant (%)

	Country of birth					
Nationality	**UK**	**A8**	**EU15**	**Non-EU**	**Pre-2004**	**Total**
UK	95	0	0	0	4	100
A8	0	99	0	1	0	100
EU15	4	0	79	17	0	100
Non-EU	4	0	1	94	0	100
Pre-2004	0	0	0	0	100	100
Total	89	2	0	2	7	100

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. Sample A, with a larger sample size due to the inclusion of non-EU nationals and pre-2004 migrants. n= 328,428.

Table A2: Origin of EU nationals born in non-EU countries (%)

Continent of birth	**% of group**
Asia	33
Africa	46
The Americas	17
Other	4
Total	100

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. n=273.

Appendix B: National origins of A8 and EU15 migrants

Table A3: National origins of A8 sample

Nationality	% of group
Czech republic	3
Czechoslovakia	0
Estonia	0
Hungary	4
Latvia	4
Lithuania	10
Poland	70
Slovakia	8
Slovenia	0
Total	100

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. Respondents in the LFS occasionally report 'Czechoslovak' nationality, hence its inclusion here, though the numbers are small and hence round to zero. n= 5,174.

Table A4: National origins of EU15 sample

Nationality	% of group
Austria	2
Belgium	3
Denmark	2
Finland	1
France	17
Germany	13
Greece	4
Ireland	14
Italy	14
Luxemburg	0
Netherlands	5
Portugal	13
Spain	8
Sweden	4
Total	100

Source: LFS. Notes: Employed men and women, aged 16-64, not in full-time education. n=1,600.

Appendix C: PISA mapping of ISCED levels to age left full-time education

Table A5: Mapping of ISCED to age left full-time education

	ISCED levels					
Nationality	**Primary**	**Lower secondary**	**(Upper) secondary**	**Post-secondary, non-tertiary**	**Vocational tertiary**	**Academic tertiary**
Austria	10	15	18	19	21	23
Belgium	12	15	18	18	21	23
Czech Rep.	11	15	17	19	22	22
Denmark	13	16	19	19	22	24
Estonia	11	16	19	19	22	23
Finland	13	16	19	19	22	24
France	11	15	18	18	20	21
Germany	10	16	19	19	21	24
Greece	12	15	18	18	21	23
Hungary	11	15	18	19	21	24
Ireland	10	13	16	16	18	20
Italy	11	14	18	19	22	23
Latvia	10	15	18	18	23	23
Lithuania	10	15	18	18	22	23
Luxembourg	12	15	18	19	22	23
Netherlands	12	16	-	18	-	22
Poland	-	15	18	19	22	23
Portugal	12	15	18	18	21	23
Slovak Rep.	11	15	18	18	20	24
Slovenia	10	14	17	18	21	22
Spain	11	14	16	18	19	23
Sweden	13	16	19	19	21	23
UK	11	14	17	18	20	21

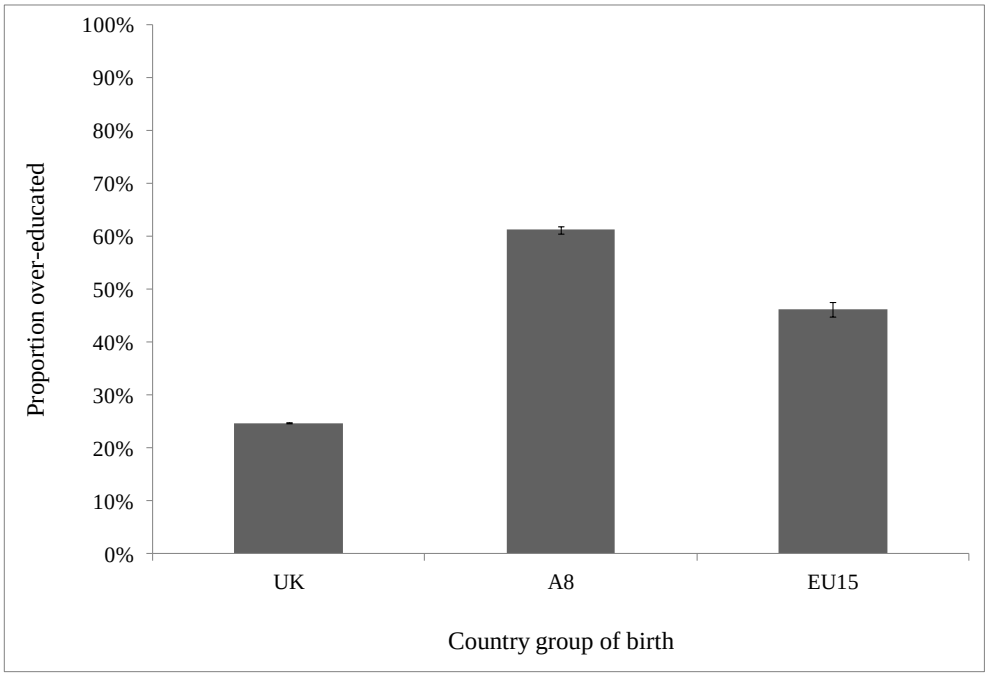
Source: OECD (2012). Notes: Czechoslovakia is set equal to the Czech Republic.

A drawback of assigning ISCED levels based on 'age left full-time education' is that this age is sometimes the same for different levels of qualification. With a conservative approach in mind, I have assumed in every case that a person has achieved the lower ISCED level, if they leave school at an age that could indicate two different ISCED levels. For example, in Lithuania, someone leaving school at 18 could either have achieved 'Upper secondary' or 'Post-secondary non-tertiary' education, but I assume they achieved 'Upper secondary'. Likewise, In the Czech Republic, someone leaving education at 22 could either have achieved 'Vocational tertiary' or 'Academic tertiary', but I assume they achieved 'Vocational tertiary'. Note that this problem does not affect Poland, which is by far the largest nationality group within the A8 sample. I assume that assigning the lower level of education here reduces the level of 'over-education' appearing in both the EU15 and A8 groups.

Appendix D: Proportion over-educated using 'Country of birth' definition

If migrants were defined by 'Country of Birth' rather than nationality in this paper, the proportions classed as over-educated would not change substantially. The difference is less than a percentage point for each group.

Figure A1: Proportion over-educated using 'Country of birth' definition of migrant, by nationality group (%)



Source: LFS. Notes: Each proportion is a mean value, and bars represent standard error of mean. Employed men and women, aged 16-64, not in full-time education. Sample A in Table 5, with a slightly smaller sample size due the different definitions of the groups. n=299,255.

Appendix E: Additional wage equations

Table A6: Log wage equations: all non-graduate occupations

	Additional control variables			
Characteristics	**Year/quarter**	**Age/gender**	**Location**	**Job type**
Mismatch				
Over-educated	22	27	25	21
	(0)	(0)	(0)	(0)
Nationality				
(Comp: UK)				
A8	-33	-30	-30	-24
	(2)	(2)	(2)	(2)
EU15	-20	-19	-25	-18
	(3)	(4)	(3)	(3)
Over*A8	-13	-17	-16	-14
	(2)	(2)	(2)	(2)
Over*EU15	16	14	13	11
	(4)	(4)	(4)	(4)
Gender				
Female		-22	-21	-13
		(0)	(0)	(0)
Female*A8		16	15	9
		(2)	(2)	(2)
Female*EU15		10	10	4
		(4)	(4)	(3)
Age				
(Comp: 16-25)				
26-30		28	27	24
		(0)	(0)	(0)
31-35		38	37	34
		(0)	(0)	(0)
36-45		43	43	38
		(0)	(0)	(0)
46-64		41	40	37
		(0)	(0)	(0)
Place of work				
(Comp: London)				
South-east			-22	-20
			(1)	(0)
Regions			-29	-26
			(0)	(0)
Job				
Part-time				-17
				(0)
Supervisor				24
				(0)
Constant term	202	177	204	197
	(1)	(1)	(1)	(1)
n	157,194	157,194	157,066	157,066

Source: LFS. Notes: Source: LFS. Notes: Standard errors in parentheses. Coefficients and SEs are multiplied by one hundred and rounded to nearest integer. Employed men and women, aged 16-64, not in full-time education, in all non-graduate occupations. Sample C in Table 5.

Appendix F: Robustness checks

Allowing for measurement error in assigning ISCED levels to EU15 nationals

As there is likely to be some measurement error caused by grade retention in the EU15 group, I have repeated the central parts of the analysis, making the assumption that any EU15 respondent who I have assigned to '(Upper) secondary', 'Post- secondary, non-tertiary' or 'Vocational tertiary', based on the age at which they left full-time education, has achieved only the '(Upper) secondary' level of education. This is a stringent assumption, but gives a lower bound for the proportion of over-education, and will highlight any potentially spurious results in the main body of the paper.

The proportion of EU15 nationals classed as over-educated using this more restrictive definition falls from 46% to 42%. Repeating the probit analysis shows the increased risk of over-education for EU15 nationals falls from 10 percentage points to 6 percentage points, with observed characteristics held constant (the effect is statistically well determined). My central analysis is therefore not substantially affected by these measurement problems, though the over-education estimates I report for EU15 nationals in the main body of the paper are potentially biased upwards. I am happy to supply these additional results tables on request.