

Accepted Manuscript

Emigrant Selection and Wages: the Case of Poland

Anna Rosso

PII: S0927-5371(19)30063-6
DOI: <https://doi.org/10.1016/j.labeco.2019.06.008>
Reference: LABECO 1744

To appear in: *Labour Economics*

Received date: 14 September 2018
Revised date: 22 May 2019
Accepted date: 26 June 2019

Please cite this article as: Anna Rosso, Emigrant Selection and Wages: the Case of Poland, *Labour Economics* (2019), doi: <https://doi.org/10.1016/j.labeco.2019.06.008>



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Emigrant Selection and Wages: the Case of Poland*

Anna Rosso[†]

June 29, 2019

Abstract

In this paper, I use a unique individual-level pre-migration labour market dataset for Poland to examine emigrant selection in two major destination countries, the United Kingdom and Germany. I compare the pre-migration observable and unobservable characteristics of emigrants with those of non-emigrants in Poland. First, I find that Polish emigrants to the UK are more similarly educated to non-emigrants while being more negatively selected on residual wages. Second, emigrants to Germany are disproportionately more likely to fall in the middle of the education distribution but they are no different than non-emigrants in terms of unobservable skills. The familiar predictions of the Borjas (1987) model allow me to partially uncover the mechanism driving the selection patterns of Polish emigrants. I contribute to the migrant selection literature by providing additional evidence on how migrants respond to differences in both labour markets and migration policies across countries.

JEL-code: F22, J61, O15, D33

Keywords: International migration, selection, skill prices, EU enlargement, inequality

*I am grateful to Christian Dustmann and Ian Preston for support and guidance throughout this project. I also thank Iga Magda for her invaluable help with the data, and Massimo Anelli, Simone Bertoli, Pedro Carneiro, Giovanni Facchini, Jesus Fernandez-Huertas Moraga, Tommaso Frattini, Simon Goerlach, Luigi Minale, Hillel Rapoport, Andrea Riganti, Anna Raute, Uta Schöenberg and Jan Stuhler for their helpful comments. My gratitude also to participants of the 4th Norface Migration Network Conference "Migration: Global Development, New Frontiers", the conference on "International Labor Mobility and Inequality across Nations", the Second Workshop on the Economics of Migration at Goethe University of Frankfurt and the First Workshop on Migration and the Labour Markets in Edinburgh, for their comments and suggestions. I am very grateful to the editor and two anonymous referees of this journal, who helped us improve significantly the paper. Financial support from the Rockwool Foundation for data provision is acknowledged. I also acknowledge the support of Uta Schöenberg, for help accessing the German data. All errors are my own.

[†]University of Milan, Department of Economics and Centro Studi Luca d'Agliano; email: anna.rosso@unimi.it

1 Introduction

Immigration is not a random process; rather, immigrant selection depends on the conditions in both the origin and destination countries. When Poland joined the European Union in 2004, Polish workers were allowed to migrate without policy restrictions to some European countries, including the United Kingdom. Germany, in contrast, opted for a 7-year transitional period before opening its borders, so Polish migration was still partly regulated by stricter flow controls. Germany and the UK could be considered major destination countries for Polish emigrants, accounting for about 49 percent of the emigrant flow from Poland in 2007 (Dustmann et al. 2015). Given the importance of Polish emigration between 1998 and 2008, a better understanding of migrant selection is crucial because of its possible consequences, affecting both the inequality level in sending countries and immigrant labour market integration in the destination countries.¹

The study of (unobservable) selection of migrants began with Borjas (1987) and was extended to the case of both observable *and* unobservable selection in Borjas (1991). The Borjas (1987) model showed that if the return to skills is higher in the origin than in the destination country, the emigration probability is higher for lower skilled than for higher skilled workers, and vice versa. Individuals with below (above) average skills may benefit more from moving to a country with lower (higher) levels of inequality. The model implies that across-country inequality differentials affect emigrant skill levels.

This paper examines the self-selection of emigrants from Poland within the EU, in particular those migrating to the United Kingdom and Germany between 1998 and 2008. I draw on a unique individual-level labour market dataset from Poland, which provides both pre-immigration labour market characteristics and final destination country, to present fresh evidence on the selection patterns across the two destination countries, and the factors that drive them. I use education to proxy observable skills, and I estimate unobservable factors based on residual wages from wage regressions for a sample of individuals, including both emigrants and those who stay. This information is used to provide an extensive description of the observable and unobservable skills of emigrants to both destination countries. I then test for the familiar predictions of the Borjas (1987) simple framework to understand the mechanism driving the described data patterns, using the same empirical strategy as Gould and Moav (2016), and measure the role of across-country inequality. I estimate the probability to emigrate as a function of skills and of skill prices (proxies for inequality) in both source and destination countries. I determine the relative prices of education and residual wages, respectively, by calculating the difference in return to education and the

¹See for example Dustmann et al. (2015) on the effect of emigration on wages of Polish stayers and also Rosso (2016), Clark and Drinkwater (2008), Drinkwater et al. (2009) and Blanchflower and Shadforth (2009) on the integration of Eastern Europeans in the UK labour market.

inherent levels of residual inequality between the destination and source countries. The latter is measured by the standard deviation, as in Juhn et al. (1993), Gould (2002) and Gould and Moav (2016), and high residual wage inequality corresponds to high rewards to unobserved skills. Hence, whereas the relative skill prices explain differences in observable and unobservable composition between emigrants and non-emigrants, the average wage differences between destination and source countries explain the differences in the average probability of migrating to that country (a scale effect).

Overall, my analysis makes a twofold contribution to the literature on migrant selection. Firstly, there is very little research focused on European migration, especially within the EU. Most of the literature has over-concentrated on Mexican migrants on their path to the U.S. Particularly valuable is my focus on a country not yet studied but which experienced high emigration rates over the period analysed. Differently from Borjas et al. (2019) who focus on emigration from a rich EU country (Denmark), Poland is to be considered the largest emerging country within the European context. Before the 2004 accession, average monthly earnings were 4 times lower in Poland than the average in the EU15 countries² and, typical of transition economies, the country experienced large fluctuations in the unemployment rate, from 10 percent in 1998 to 19 percent in 2004, and back to 7 percent in 2008.³ Simultaneously, the emigration rate increased from 0.5 percent to 2.17 percent Dustmann et al. (2015), making Poland a very interesting case study. My second contribution broadens the scope of the emigrant selection literature by assessing, in a new scenario, the role of differential levels of inequality of both observable and unobservable skills as an economic incentive to migrate. I implement the empirical analysis of Gould and Moav (2016) for this case and extend it to the comparison of two destination countries, reviewing the role of migration policies as possible drivers of the observed outcomes. As argued by Borjas et al. (2019) there is a need for further evidence to improve the understanding of the role of residual wage inequality in determining emigrant selection. The literature has so far pointed out the importance of focusing not only on observable skills but also on unobservable skills, allowing for the possibility that individuals with identical education levels have different levels of unobservable skills (Gould and Moav 2016; Borjas et al. 2019). The theoretical model is applicable only to individuals who are employed, nevertheless, my dataset allows me to analyze the characteristics of individuals who were unemployed before they emigrated. I therefore provide a novel contribution on the descriptives of this interesting sample, showing how employed and unemployed emigrants are very similar across most characteristics.

I find that Polish emigrants to the UK are more similar to non-emigrants with regards to

²Eurostat data on Structure of Earnings Survey, <https://ec.europa.eu/eurostat/web/microdata/structure-of-earnings-survey>

³The average unemployment rate between 1998 and 2008, in EU15 countries was 8-9 percent.

education, while Polish emigrants to Germany are disproportionately more likely to fall into the middle of the distribution of education. With regards to unobservable characteristics, Poles in the UK are negatively selected, while emigrants to Germany have no different residual wages than the underlying Polish population. These data patterns partially support the prediction of the model. In particular, I find that the probability to emigrate decrease with residual wages for emigrants to the UK, consistent with the evidence of higher residual inequality in Poland than in the UK. Yet, I observe that residual wages are not positively correlated with the probability to emigrate for emigrants to Germany, despite residual inequality being lower in Poland than in Germany. Finally, in no cases does the probability to emigrate decreases with education, a result the model would predict given the higher return to education in Poland versus the UK and Germany; however, this finding confirms most of the results on observable selection uncovered in the literature. Following Gould and Moav (2016), I test the model further to check for systematic variation across occupation in the returns to both observable and unobservable skills that would bring more insights to the results, which are nevertheless confirmed also within occupation. I find in fact that the lower return to unobservable skills in the UK versus Poland, within occupational sectors, creates an incentive for Polish individuals with below-average residual wages to emigrate to the UK, while the higher unobservable returns in Germany versus Poland do not entice Poles with above-average residuals to leave. Finally, a higher return to education in the home country compared to the destination does not induce low-educated Poles to leave the country. This result confirms the findings of Gould and Moav (2016). Yet, I further contribute to the literature by taking an additional step and looking at the return to education of a different reference group in the destination country, the Polish immigrants. In this case, results are more supportive of the model predictions even within occupations.

Additional implications of the Borjas model that are confirmed for the UK but not for Germany highlight the importance of other confounding factors (networks, migration policies, etc.) in defining migrant selection patterns. That is, existing networks, policies and visa requirements can interfere with the selection process by influencing both the size and characteristics of migration (Abramitzky et al. 2012; Gould and Moav 2016). For instance, after migration restrictions in the UK were abolished in May 2004, emigrants from Poland were free to work in the country without any extra requirements. Under these conditions, migration flows occurred solely as a consequence of individual economic decisions, while in Germany, selection was probably still driven by migration policies (see section below). Interestingly, these results may also imply that migration flows are determined by factors that remain unobserved by immigration officers but are crucial in defining migration quality and thus relevant for policy making (Bertoli et al. 2016). Therefore, the UK versus Germany comparison allows me to qualitatively compare selection in a situation with no

barriers to migration, to one with a more policy-regulated scenario.

The advantage of using pre-migration data is its ability to provide a full picture of the bases for emigrant selection. In fact, most studies that use pre-migration information designate unobservable skills and previous earnings as major components of the selection process. Indeed, Kaestner and Malamud (2013), Fernández-Huertas Moraga (2013), and Ambrosini and Peri (2012) all show that emigrants from Mexico fall into the middle of the education distribution but are more negatively selected in terms of earnings. This negative selection can partly be explained by differential returns to labour market skills: not only is a mere 5 percent of the earnings gap explainable by observable characteristics, but most selection happens on unobserved characteristics (Ambrosini and Peri 2012). In closely related work, Gould and Moav (2016), who use pre-migration labour market data for Israel to measure observable and unobservable selection for a set of emigrants assumed to move to the U.S., demonstrate that the emigration probability increases with education whereas unobservable skills exhibit an inverse U-shaped pattern. These authors thus claim that unobservable skills can be decomposed into general and country specific, and that above certain levels the benefits of migration decrease. Overall, Borjas et al. (2019) emphasize that any analysis of migrant earnings based on observable characteristics alone can understate the actual selection, because of the major role played by unobservable characteristics. To support this claim, they show that in a sample of emigrants to Denmark, more than half of the differences in expected earnings between movers and stayers can be ascribed to differences in unobserved characteristics.⁴ Within the same context as my study (the EU), this most recent contribution studies migrant selection from a more egalitarian country to other European countries and to the United States. The empirical results are in fact in line with the predictions of the model, finding that the earnings distribution of emigrants stochastically dominates that of the individuals who stay in Denmark.⁵ I contribute to the literature by demonstrating how the model predictions also apply to a less egalitarian country within the European Union, where unobservable characteristics also play a major role in selection.

The remainder of my empirical investigation unfolds as follows: Section 2 describes the migration policies in place in the two destination countries; Section 3 introduces the data and presents

⁴Parey et al. (2017) measure the selection of highly skilled emigrants from Germany by comparing the predicted earnings of migrants versus non-migrants. The selection of university graduates is consistent with the predictions of the Borjas model; in fact, migrants to less equal countries have higher predicted earnings, while migrants to more equal countries have lower predicted earnings. Using alternative measures of skills, see also Patt et al. (2017) where occupational skills (manual vs cognitive) capture economic incentives to emigrate from Mexico.

⁵Evidence supporting the Borjas (1987) model predictions has emerged in contexts where migration policies were not binding and selection was driven solely by economic incentives. Norwegian migrants, during their historical mass migration to the U.S., were more negatively selected than Norwegian non-emigrants in a period when the Norwegian income distribution was more unequal than that of the U.S. (Abramitzky et al. 2012). At this time, migration policies were non-existent and so did not interfere with individual decisions. Additional evidence that supports the migrant selection model of Borjas is also provided by studies of internal migration as individuals can move without restrictions (Borjas et al. 1992).

the observed patterns of selection for the two destination countries; Sections 4 and 5 investigate the possible mechanisms that could drive the observed patterns as well as giving an interpretation of the results and some extensions; Section 6 concludes the paper.

2 Polish immigration to the UK and Germany

Polish migration to the UK began after WWII, with the British government favouring the resettlement of Polish Armed Forces (and their families) and Polish refugees from the European labour camps. Nevertheless, despite this programme, the flow of Polish immigrants was unremarkable and involved primarily political dissidents (Trevena 2009). Subsequent economic migration, on the other hand, was regulated by the 1971 Immigration Act, which subjected Polish workers in the UK to the same rules applied to non-EU nationals, meaning that they needed a work permit, issued for a limited time and by a specific employer. On the other side of the channel, before the 1990s, most migration between Poland and Germany involved ethnic Germans (Eastern European *Aussiedler*), 3 million of whom -50 percent of them Polish (Fassmann and Munz 1994)- entered the country between 1950 and 1993. At that time, German law allowed people of German ethnicity living in countries of the former Warsaw Pact, including Yugoslavia, to return to Germany and claim German citizenship. These rules changed in 1992, however, giving only former residents of the Soviet Union this right, and allowing only Polish *Aussiedler* to stay in Germany via permanent visa or family reunion. However, following the collapse of the Communist regimes in Eastern Europe, and German reunification, the increase in German labour demand forced the government to implement new entry schemes to direct immigrant flows and create employment opportunities primarily for Eastern European citizens. Eastern European programme workers were mainly concentrated in low-level jobs, and more than 80 percent of actual labour from Poland, the Czech and Slovak Republics and Hungary came as guest workers and were allowed to stay for up to 18 months (Honekopp 1997).⁶

⁶There were five different programmes: 1. Project-tied worker schemes: a German company could subcontract part of a project to a foreign firm, so workers could stay in Germany for the duration of the contract. Because they had to be socially insured in their home country, the hiring costs to the German companies were low. This type of contract was primarily used in the construction sector but was stopped in 1997 after legal intervention of the European Commission. 2. Seasonal-worker scheme: workers could stay in Germany for up to three months if no German workers were found for the position. The employer was supposed to apply for this scheme, and the seasonal contracts were used in the farming and hotel/restaurant sectors. 3. Border commuters: Polish and Czech nationals could be hired if they lived within 50 km of the border and if no German worker could be found for the post. However, they were not granted residence and were forced to commute on a daily basis. 4. Guest workers: this scheme was targeted at young workers from Eastern European countries (aged 18 to 40) who could move to Germany to improve their professional skills. It had educational requirements (at least vocational training) and a basic knowledge of the language. The number of visas issued through this programme was restricted to 1,000 a year for Polish workers. 5. Foreign nurses: This scheme was created to meet the demand for nurses and mainly attracted immigrants from the former Yugoslavia (Honekopp 1997).

After 1989, negotiations to enter the EU began in 10 Central and Eastern Europe countries, among them Poland.⁷ The first relevant change happened in December 1991 when the EU signed Association Agreements (ratified in February 1994) with Poland and Hungary and later (1993-1996) with other candidate countries. These early agreements aimed at progressively establishing bilateral associations between the EU and candidate countries on several community topics, including the movement of workers. In fact, at that time, although host Member States gave legally employed nationals from candidate states the same rights as other workers, and granted special labour market access to their spouses and children, they gave them no right of access to the labour market without permits or freedom to reside in the host country, thereby extending no social protection. Nevertheless, these agreements could vary by candidate state, with Polish nationals, for example, allowed to receive work permits if they already had residence permits in other Member States. As a consequence of these new Association Agreements, Eastern European nationals were also given the freedom to establish businesses in one of the Member States, with the pre-2004 exceptions of financial services, real estate and privatisation operations. Hence, for the 10 years prior to 2004, Polish immigrants to both the UK and Germany could enter the EU via the self-employment route.⁸

Then in May, 2004, Polish immigrants gained the right to access the British labour market with no restrictions, which implied that they could enter the country and take up employment without a work permit. The only other two countries that completely lifted all the barriers to migration from new accession countries were Sweden and Ireland. The rest of the EU, including Germany, opted for a transitional period, lasting up to 7 years, during which the free movement of workers was still regulated by the system in place at the time of the enlargement. Hence, during these 7 years, with the exception of self-employment as regulated by the 1991 Agreement, Polish immigrants to Germany still needed an employment contract to obtain a work permit.⁹

⁷Together with Poland, seven other Eastern Europe countries entered in 2004 (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Slovakia and Slovenia), as well as Malta and Cyprus. Romania and Bulgaria entered in 2007.

⁸Many used this route as there were no capital requirements to set up a business and they could apply to this scheme after arriving in the destination country. The illegal route to the UK was likewise an option, as many entered the country with tourist and student visas, then remained (Trevena 2009). Similar trends, together with temporary migration and commuting, were also observed in Germany (Miera 2008).

⁹Some privileged routes were already in place during this period, and citizens of new accession countries, for example, still had a priority over non-EU citizens when offered a job.

3 Data and Statistics

3.1 Data sources and sample selection

The study data are taken from the 1998-2008 waves of the Polish Labour Force Survey (PLFS), a rotating quarterly panel of about 15,000 households (50,000 individuals) per quarter, administered by the Polish Central Statistical Office (GUS) in all of Poland's 16 provinces (*voivodeships*). The entire interview period spans 1.5 years, with each household interviewed four times and each individual interviewed in two consecutive quarters, and then in another two consecutive quarters following a gap of two quarters. The survey covers all individuals aged 15 and above within a sampled household. These PLFS data contain the demographic, personal and household characteristics of all interviewed individuals, including age, education, current and past region of residence (within Poland), country of birth, number of children and the economic activity of each household member during the week preceding the interview. The survey also collects information on household members who have been residing abroad for at least 3 months (whom I define as emigrants), which implies that it excludes short-term and seasonal workers. Given the previously described German legal restrictions for seasonal migrants, this dataset allows me to capture longer-term migration.¹⁰

Because the survey is rotational, I observe a sub-sample of emigrants originally resident in Poland who decided to emigrate during the sampling period. For these individuals, I not only have PLFS information on age, education, country of emigration and role within the household, but also pre-emigration labour market information -including, monthly wages (in zloty) and occupation- which can be merged with their emigrant status and, most importantly, with their final destination. By averaging pre-emigration real monthly wages across all periods before they emigrate, for all emigrants¹¹, I construct a measure of emigrant wages and hours worked. In doing so, I do not rely only on the individual wage information just before migration, so as to avoid capturing any negative shocks suffered immediately previous to the migration decision. The final emigrant sample is restricted to individuals aged 16 to 64 who were employed before migrating, and the non-emigrant sample to employed individuals aged 16 to 64 with no missing wage data, who were living in Poland over the same period. Using emigrant information from the source country dataset has two distinct advantages: first, it allows a more precise comparison between emigrants and non-emigrants on observable characteristics like education, which can be misleading across data sources from different countries; second, if the undercounting of immigrants in the surveys from the

¹⁰The survey also asks how long the individual has been abroad and for what reason: 50 percent of emigrants have been away more than 1 year, over 85 percent for work-related reasons.

¹¹Taking the average wage over time instead of the very last information before they migrate allows me to have a measure of wages which is less endogenous to the migration decision, and also better captures any dynamics in the skill development of the individual over time. Nevertheless, this restriction is not the driver of the results.

destination countries disproportionately depends on immigrant characteristics, their inclusion in the destination country sample can be correlated with their legal or labour market status and hence their skills.¹² The greatest benefit of using this dataset, however, is the ability to define selection in terms of both observable and unobservable characteristics, with the latter measured by residual wages. Unobservable selection is then a measure of selection within observable characteristics, one that allows me to rank individuals within occupation, education and region. On the other hand, it should be noted that this type of selection is only definable if the individual is employed prior to migration. In this respect, the decisions to leave the country taken by individuals who are not employed, or are outside the labour market could be driven by different factors.

Nevertheless, because the omitted wage information can result from either sampling or labour market status, in Appendix B I use observable characteristics to determine whether the sample of emigrants used in the analysis is comparable to the sample of all emigrants included in the survey (including those with no wage data). Results show that when only observable characteristics are used, there are no differences between these two samples.

Yet despite these several advantages, using this dataset does have one drawback: it does not include emigrants living in households whose members have all emigrated.¹³ This potential undercounting of emigrants could affect the selection results if the individuals from these emigrant households had statistically significant different characteristics from those of the emigrants observed in the data. However, as Dustmann et al. (2015) point out, the share of those living in single households in Poland is low compared to that in other countries, and more frequent among the elderly. I look at emigrants in the age group 16 to 64, therefore my results are less likely to be affected by this underreporting. Polish family migration to the UK is also reportedly low (Drinkwater et al. 2009). Moreover, data from the UK Labour Force Survey, show no observable selection across households types. In terms of education, Polish immigrants whose household is formed by at least two or three Polish-born individuals are statistically the same as Polish immigrants who live alone (or with no other Polish-born individuals). In Germany, single Polish immigrants with no children, independently of the year of arrival, are statistically the same in terms of education with respect to Polish immigrants with children. These results confirm that across different types of households in the receiving countries, Polish workers have similar levels of education. In order to assess the effect that missing households could have on unobservables,

¹²Data from the source country strongly contribute to the understanding of selection among Mexican emigrants to the U.S., especially given the high rates of illegal immigrants not usually represented in destination country surveys. Fernández-Huertas Moraga (2010), on finding more negative selection among this population when using source country data, argues that previously reported results are driven by the use of destination country data. In earlier work, Ibarra and Lubotsky (2007) show that Mexican immigrants to the U.S. tend to over-report their educational level. Hence, given that illegal migration to both the UK and Germany prior to 2004 has been documented (Miera 2008; Duvell 2004), using destination country data could raise similar selection issues.

¹³See Dustmann et al. (2015) for a more detailed discussion.

I perform an exercise in a similar fashion to Lee (2009). I estimate the lower and upper bounds within which the average residual wages for individuals in single households may lie, under the following assumption: the share of single households in the underlying Polish population is the same as the share in the emigrant population (missing by construction). In this setting, sample selection is in fact caused by emigration. I also expect that the non-emigrant sample does not suffer from any attrition. I compute the share of single households by year and region, which is on average 8 percent. The share of individuals living in single households represents the share of missing wages for emigrants. This share is therefore the trimming proportion used to compute the bounds. Through this trimming exercise, upper and lower bounds of the average residual wages are estimated. The results are reported in Table C.1 in Appendix C. For emigrants to the UK, the bounds within which unobservable skills lie still imply negative selection. By contrast, the lower bound for residual wages of emigrants to Germany would suggest some negative selection.

Other data sources

The decision to emigrate depends on both the differences in average wages and on the differences in inequality (proxied by returns to skills) between home and host countries. My analysis therefore not only considers wages but also the estimated returns to observable and unobservable skills in the UK and Germany. As detailed in Appendix A, for the United Kingdom, I use 1998-2008 data from the UK Labour Force Survey (hereafter, UKLFS), which has collected information on demographic and labour market characteristics (net and gross wages) from a sample of private households since the mid-1970s.

For Germany, I use the 1998-2008 data from the German Microcensus, administered to a representative sample of around 800,000 individuals every year, which collects information on demographic and primary labour market characteristics. The microcensus also reports net monthly income. In the analysis, I weight all data using the population weights included in each dataset.

Sample selection and variables construction

I restrict the analysis to working age individuals (16-64) who report being employed, and exclude the self-employed for whom wage information is only reported in the German Microcensus. I measure wages based on net hourly wages (converted into constant 2008 zloty). In contrast to monthly wages, hourly wages should be less affected by the decision to participate in the labour market, so the inclusion of men and women together in the sample makes selection due to different labour market attachments less relevant.¹⁴ Net rather than gross wages is a choice well-supported in the literature: first, varying tax schedules across countries may result in returns to skill being overstated if they are measured using pre-tax wages (Alesina and Angeletos 2005); second, because

¹⁴Using hourly wages may not completely solve the selection problem in that part-time jobs, for example, are systematically paid less than full-time jobs and more popular among female workers (Manning and Petrongolo 2008).

the same amount of gross income in the destination country may imply a higher or lower tax burden than in the home country, these taxation differences can imply even larger across-country wage gaps (Bertoli et al. 2013); lastly, emigrants tend to sort more strongly on post-tax than on pre-tax wages (Grogger and Hanson 2011). Nevertheless, I also show results using gross wages in the Subsection 5.1 and Table C.5.

To further facilitate across-country comparability, I define the education variable based on years of education rather than qualifications, which are difficult to compare internationally. For all three countries, I use years of education to define three categories: primary educated, those who left education at age 16 (or before); secondary educated, those who left education between 17 and 20 (22 for Germany); and tertiary educated, those who left education any time after that. I thus drop from the sample all those who were still in school at the time of interview. To standardize the occupation variable across countries, I recode occupations in all the datasets to match the International Standard Classification of Occupations (ISCO-88) codes. Specifically, using the first digit of the codes, I define three occupational sectors: the professional sector includes all those employed in professional, managerial and technical occupations; the service sector includes all service, clerical and sales workers; and the blue-collar sector includes agricultural workers, craftsmen and trade workers, machine operators and labourers.

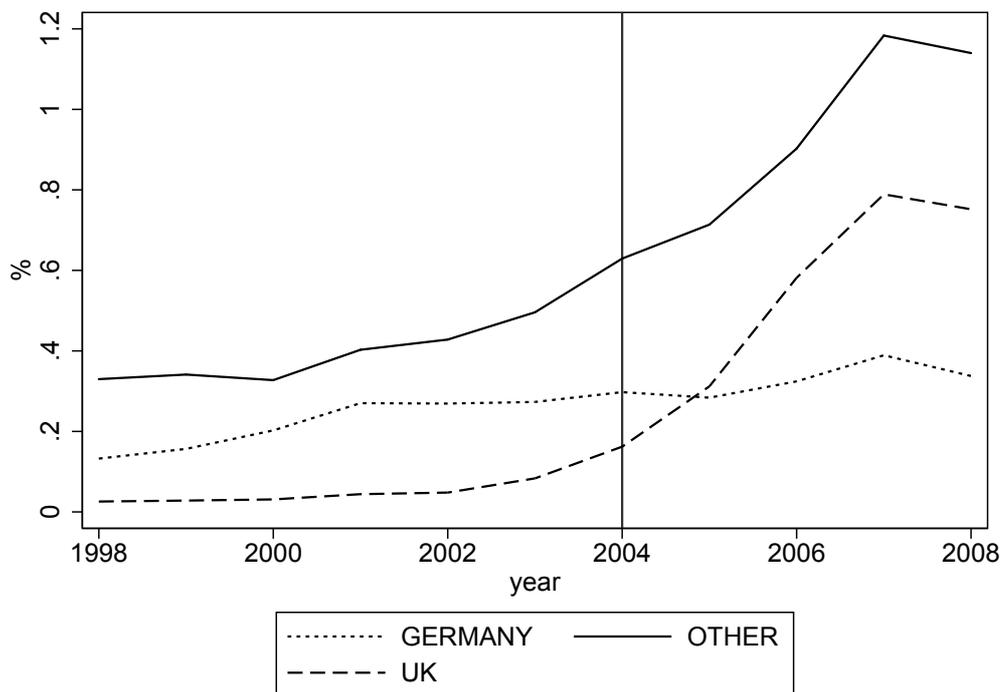
3.2 Descriptive statistics

As Figure 1 shows, the share of emigrants among the total resident population has been increasing rapidly since 2004, particularly in the UK (from 0.03 percent to 0.75 percent) and other countries (from 0.32 to 1.13 percent)¹⁵, although the increase in the share to Germany has been milder (from 0.13 and 0.33 percent). The evolution of this increase by education for the UK and Germany is shown in Figure 2, which gives the share for each destination as $\frac{Emigrants_{jt}^{dest}}{Non-Emigrants_{jt}}$, where j is educational level and t is year. As is apparent, over the years, the two countries have attracted migrants with different education levels: whereas the UK has always had the highest share of highly educated migrants, Germany has always had the highest share of the intermediate educated, and both show a low share of the less educated, albeit slightly higher in Germany. In addition, whereas EU enlargement triggered a large increase in UK shares of both intermediate and highly skilled migrants, Germany's shares of these have remained fairly constant, with the share of highly skilled even decreasing after 2004.

Table 1 lists the means for log hourly wages (in 2008 prices) and other characteristics used throughout the analysis (education, age and occupation) for non-emigrants (column 1) and for

¹⁵Other countries include European and extra-European countries, mainly: Ireland, USA, Italy, France, Netherlands, Norway, Sweden, Spain, and Belgium.

Figure 1: Share of Polish emigrants by destination country



Source: Polish Labour Force Survey. This graph shows the share of Polish emigrants by destination country. The share is computed as the number of Polish emigrants to a destination divided by the total resident population ($\frac{Emigrants_{jt}}{Non-Emigrants_{jt}}$). All emigrants available in the survey are included.

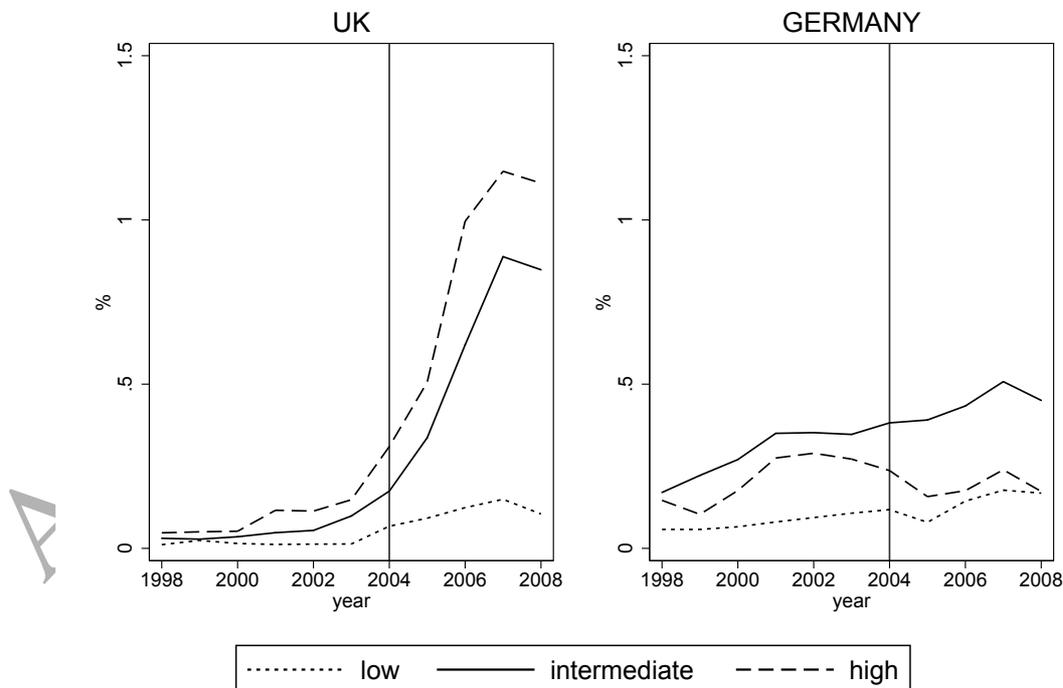
emigrants (columns 2 to 5) averaged over the 1998-2008 period. I further divide emigrants by destination country.¹⁶ The data show that although emigrants overall earn less on average than non-emigrants, emigrants to the UK have lower wages than emigrants to Germany, with emigrants to other destinations ranking somewhere in between. More specifically, the unconditional gap between emigrant and non-emigrant log wages is 16 log points (15 percent) for those to the UK and 9 log points (8 percent) for those to Germany, although these differences could admittedly result from differences in characteristics. On the other hand, the wage gap between non-emigrants and UK emigrants is surprising, especially given their similarity with regard to education: 5 percent (versus 8 percent) of UK emigrants are primary educated, 73 percent (versus 70 percent) are secondary educated and 21 percent (versus 22 percent) are tertiary educated. UK emigrants also stayed in school three months longer, but this difference is not statistically significant ($SE = 0.19$). In Germany, by contrast, whereas the share of primary educated emigrants matches that in the overall worker population (8 percent), the share with intermediate education is much larger (83 percent) and the share of tertiary educated is less than half that of non-emigrants. Overall, German emigrants have 8 months less schooling than non-emigrants ($SE = 0.153$) and a year

¹⁶For completeness and descriptive purposes, I also report emigrants to other destinations, which are then dropped from the analysis.

less than UK emigrants ($SE = 0.24$). In addition, whereas the migration probability increases with education for emigrants to the UK, it is highest for the intermediate educated emigrants to Germany (see Figure C.3).

With regard to age, although migrant workers are in general younger than those in the overall worker population, some differences are observable across destinations. UK migrants, 29 years old on average, with over 80 percent in the 16-35 age group, are the youngest, whereas German migrants are 7 years older (36) but more similar to non-emigrants who are, on average, 38 years old. The data also show that men are more likely to emigrate, with an overall share of 75 percent (see Table 1, column 2), 71 percent for the UK and 80 percent for Germany. Finally, the bottom part of the table shows the distribution of occupations, which, as I do for wages, I measure using the mode of all pre-migration observations rather than the most recent, in order to avoid capturing any negative pre-departure shocks. As Table 1, column 1 shows, professionals and blue collar workers are the most popular groups among non-emigrants, at 32 and 46 percent of workers, respectively. In contrast, emigrants are underrepresented among professionals across all destination countries, but particularly in Germany (8 percent only). Whereas the German emigrants were very likely to be blue collar (76 percent) rather than service workers (17 percent), 26 percent of emigrants

Figure 2: Share of Polish emigrants in the UK and Germany by education level



Source: Polish Labour Force Survey. The share is computed as the number of emigrants to a destination by the total resident population ($\frac{Emigrants_{jt}^{dest}}{Non-Emigrants_{jt}}$). Primary educated are those who left education at age 16 (or less), secondary educated are those who left education between 17 and 20 years old and tertiary educated are those who left education any time after that age. All emigrants available in the survey are included.

to the UK were working in the service sector and 62 percent in blue collar jobs. In fact, the occupational distribution of emigrants to the UK is slightly more similar to that of the total emigrant population, and while all the other differences in shares are statistically significant, the 4 percentage point difference in service worker share between UK emigrants and non-emigrants is not.

Overall, these statistics suggest two conclusions. First, blue collar jobs are the most popular occupation types among both emigrants to the UK and Germany, although, as Table 6 shows, this observation is hard to interpret as negative selection because the average wages in this sector are higher than average wages in the service sector, not only in Poland but in both destination countries. Second, although emigrants to the UK have the highest educational levels, they are younger than emigrants to Germany and may be more likely to work as service workers.

In Figure C.1, I compare the unconditional distribution of log hourly wages pooled over the 1998-2008 period for the non-emigrant population, emigrants to Germany and emigrants to the UK. Graphs A and B report the probability density functions and cumulative distribution, respectively. To determine whether the separate emigrant wage samples for Germany and the UK fall into the same distribution as the non-emigrant wage sample, I perform a two-sample Kolmogorov-Smirnov test, which yields D statistics of 0.193 and 0.108 for emigrants to the UK and Germany, respectively. I thus reject the null hypothesis of an identical underlying wage distribution for emigrants and non-emigrants at the 1 percent (for the UK) and 5 percent (for Germany) significance levels. Running the same test to check the differences in wage distribution between emigrants to the UK and those in Germany yields a D statistic of 0.122, meaning that the null hypothesis is again rejected at a 7 percent significance level.

3.3 Emigrant versus non-emigrants wages

The across-group wage differences identified above may result from differences in the regional distribution of emigrants and non-emigrants; for example, emigrants to the UK may systematically come from regions that pay lower wages, which could explain the lower values in Table 1. To investigate this possibility, I order regions based on 1998-2008 regional real log average wages, and assign them to four categories according to the percentiles of this distribution. In Table 2, for each regional group¹⁷, I report the percentage of individuals that (used to) live there (column 1), their average of log hourly wages in real terms (column 2) for non-emigrants (Panel A), emigrants to the UK (Panel B) and emigrants to Germany (Panel C). For emigrants only, I report the raw

¹⁷Each group includes the following regions: below the 25th percentile: Kuyavian-Pomeranian, Swietokrzyskie, Warmian-Masurian, Łódź and Greater Poland; between the 25th and 50th percentile: Subcarpathian, Lubusz, Lublin, West Pomeranian, Pomeranian, Opole; between the 51th and 75th percentile: Podlaskie and Lower Silesian; above 76th percentile: Lesser Poland, Silesian and Masovian.

Table 1: Individual characteristics (non-emigrants and emigrants by destination)

	Non-Emigrants	Emigrants			
	(1)	All destinations (2)	UK (3)	Germany (4)	Other (5)
log hourly wages (real)	1.97 (0.001)	1.86 (0.013)	1.81 (0.035)	1.88 (0.040)	1.86 (0.015)
Education (years)	12.53 (0.004)	12.41 (0.062)	12.77 (0.179)	11.71 (0.173)	12.43 (0.071)
Primary	8% (0.000)	7% (0.006)	5% (0.016)	8% (0.020)	7% (0.007)
Secondary	70% (0.001)	74% (0.012)	73% (0.033)	83% (0.032)	73% (0.014)
Tertiary	22% (0.001)	19% (0.011)	21% (0.030)	9% (0.027)	20% (0.012)
Age	38 (0.017)	32 (0.235)	29 (0.541)	36 (0.780)	32 (0.271)
Age 16-25	10% (0.000)	21% (0.010)	27% (0.032)	10% (0.023)	21% (0.012)
Age 25-35	29% (0.001)	47% (0.013)	56% (0.036)	38% (0.041)	47% (0.015)
Females	46% (0.001)	25% (0.012)	29% (0.033)	20% (0.035)	25% (0.013)
Professional	32% (0.001)	13% (0.009)	11% (0.024)	8% (0.026)	13% (0.010)
Services	22% (0.001)	24% (0.011)	26% (0.032)	17% (0.031)	24% (0.013)
Blue collar	46% (0.001)	63% (0.013)	62% (0.036)	76% (0.037)	62% (0.015)
Observations	410,587	1,564	204	162	1,207
Probability to emigrate		0.40%	0.05%	0.04%	0.3%

Source: Polish Labour Force Survey, years 1998 to 2008. Individuals who report a positive wage, who are employed, as there is no information on self-employed, aged 16 to 64. Emigrants are Polish citizens who are now living in another country but who lived in Poland during the former interviews of the survey. Log hourly wages in 2008 prices. Primary educated are those who left education at age 16 (or less), secondary educated are those who left education between 17 and 20 years old and tertiary educated are those who left education any time after that age. The professional sector includes all workers in the professional, managerial and technical occupations. The service sector includes all service, clerical and sales workers. The blue-collar sector includes agricultural workers, craftsmen and trade workers, machine operators and labourers. Probabilities to emigrate are computed using only the sample of individuals with positive wage information. Entries are weighted with the survey weights for non-emigrants and the estimated weights for emigrants following Dustmann et al. (2015).

difference between non-emigrant and emigrant log hourly wages (column 3). As the table shows, whereas 66 percent of the emigrants to Germany were located in regions at the bottom of the distribution (below the 50th percentile), similar to that of the stayers (64 percent), the share of

emigrants to the UK is 73 percent. In the poorer regions, emigrants to the UK earn significantly lower wages, whereas in the richest regions, their average wages are not statistically different from those of the non-emigrant population. The average wages of emigrants to Germany are lower and significantly different from those of the non-emigrant population only in the middle income regions, where emigrants are also over-represented. Hence, whereas the lower wages for emigrants to the UK (see Table 1) are partly explainable by differences in their regional distribution, emigrants to Germany are regionally distributed in a manner similar to non-emigrants. Moreover, within-regional differences are significant at the bottom of the regional distribution for emigrant to the UK.

Table 2: Differences in log hourly wages between non-emigrants and emigrants by group of regions

Regional ranking	A. Non-Emigrants		B. UK			C. Germany		
	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(3)
below 25th	32%	1.913 (0.002)	40%	1.721 (0.039)	0.192*** (0.039)	34%	1.915 (0.059)	-0.003 (0.059)
25th and 50th	32%	1.946 (0.001)	33%	1.723 (0.066)	0.223*** (0.066)	32%	1.822 (0.065)	0.124* (0.054)
50th and 75th	11%	1.972 (0.002)	10%	1.858 (0.077)	0.114 (0.078)	21%	1.769 (0.070)	0.203*** (0.073)
above 75th	25%	2.037 (0.001)	16%	2.072 (0.084)	-0.035 (0.084)	12%	2.136 (0.151)	-0.099 (0.107)

Source: Polish Labour Force Survey, years 1998 to 2008. Individuals aged 16 to 64, reporting positive wages. In each Panel (A,B,C), column (1) reports the regional distribution of individuals in the reference groups, column (2) reports the log hourly wages (in 2008 prices) between 1998 and 2008. Column (3) in panels B and C, reports the difference in log hourly wages between non-emigrants and emigrants. Regions are ranked according to the distribution of average regional log hourly wages (in 2008 prices), between 1998 and 2008, based on this distribution I create 4 groups (regions below the 25th percentile of average regional wages, between the 25th and 50th, between 50th and 75th, and above 75th). Standard errors reported in brackets. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

I extend this analysis by using information from the previous section to assess whether these emigrant and non-emigrant wage differences are explainable solely by differences in their characteristics. To answer this question, I non-parametrically construct a counterfactual distribution of emigrant wages by re-weighting the non-emigrant wage distribution with emigrant characteristics. The difference between the actual non-emigrant distribution and the counterfactual emigrant distribution thus measures selection due to observable characteristics.¹⁸ Figure 3 plots the kernel

¹⁸More details on how the exercise is performed are provided in Appendix B. The price function in this case is that of non-emigrants ($g_{non-em,PL}$), while the characteristics density is that of the emigrants ($h_{emi,PL}$). With this information I can estimate: $\hat{f}_{non-em,PL} = \int g_{non-em,PL}(w|x)h_{emi,PL}(x)dx = \int g_{non-em,PL}(w|x)h_{non-em,PL}(x)\frac{h_{emi,PL}(x)}{h_{non-em,PL}(x)}dx$, where the ratio is calculated using Bayes' rule $\frac{h_{emi,PL}(x)}{h_{non-em,PL}(x)} = \frac{\frac{Pr(emi|X)}{1-Pr(emi|X)}}{\frac{Pr(emi)}{1-Pr(emi)}}$, where $Pr(emi|X)$ is the conditional probability to emigrate estimated using a

density estimates of the non-emigrant wage (log hourly wage in 2008 prices, solid line) distribution over the 1998-2008 period against the actual density of emigrants ($f_{emi,PL}(w)$, dotted line) and the reweighted density functions ($\hat{f}_{non-emi,PL}(w)$, dashed line) for emigrants to the UK (Panel A) and emigrants to Germany (Panel B). The reweighted density functions, which correspond to the counterfactual distributions, represent what emigrants would earn if their observable characteristics (age, education, gender, region, occupation) were paid according to the non-emigrant price. The straight lines in the kernel density plots are the median wages in both the non-emigrant (solid line) and re-weighted wage distribution (dashed line). The difference in the actual wage distribution between non-emigrant and emigrant can be decomposed as follows:

$$f_{non-emi,PL}(w) - f_{emi,PL}(w) = \underbrace{f_{non-emi,PL}(w) - \hat{f}_{non-emi,PL}(w)}_{\text{observable selection}} + f(\epsilon) \quad (1)$$

Here, I define observable selection based on the difference between actual and counterfactual non-emigrant kernel density: $f_{non-emi,PL}(w) - \hat{f}_{non-emi,PL}$, while defining unobservable selection $f(\epsilon)$, as $\hat{f}_{non-emi,PL}(w) - f_{emi,PL}(w)$, so I can summarize selection using the first moments of these distributions:

$$E_{non-emi,PL}(w) - E_{emi,PL}(w) = \underbrace{E_{non-emi,PL}(w) - E_{non-emi,PL}(\hat{w})}_{\text{observable selection}} + \bar{\epsilon} \quad (2)$$

Table 3: Total wage gap decomposition

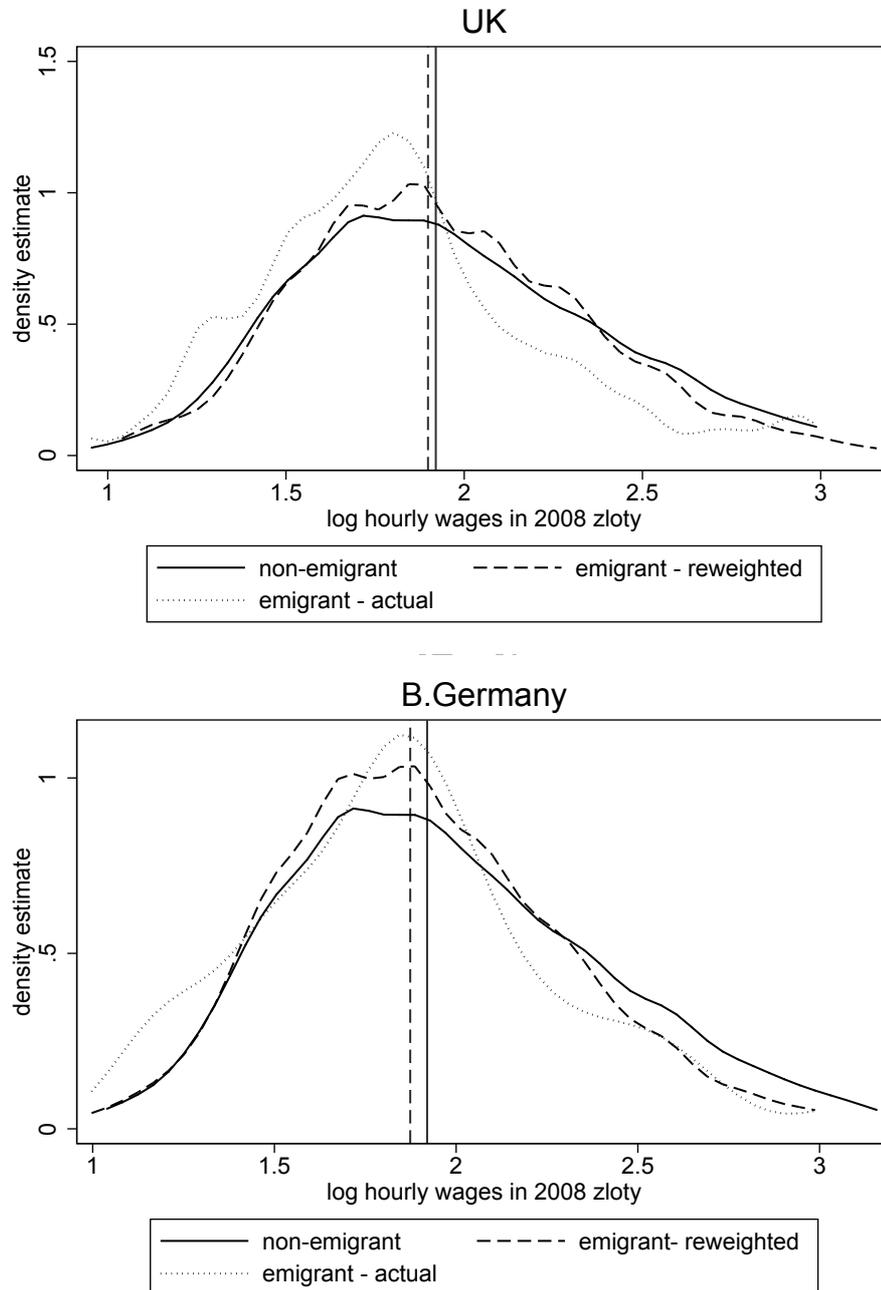
	Total wage gap	Observable selection	Unobservable selection
	(1)	(2)	(3)
UK	0.162	0.068*** (0.001)	0.093*** (0.018)
Germany	0.086	0.066*** (0.008)	0.019** (0.009)

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Average wage gaps between non-emigrants and emigrants. Observable selection is $E_{non-emi,PL}(w) - E_{non-emi,PL}(\hat{w})$ and unobservable selection is $\bar{\epsilon}$, as reported in equation 2

In Table 3 I report the results from this decomposition. Whereas the wage gap between non-emigrant wages and emigrant wages is 0.162 for emigrants to the UK and 0.086 for those to Germany (column 1), the average wage for emigrants to the UK in the re-weighted distribution

probit model of the dummy emigrant on observable characteristics, including years of education, 5 age-group dummies, region and year dummies, separately for emigrants to the UK and emigrants to Germany.

Figure 3: Kernel density estimate of log hourly real wages and differences.



Source: Polish Labour Force Survey, years 1998 to 2008. For the estimation of the kernel densities on the left-hand side plots, I use the Gaussian kernel function with bandwidth 0.06, as in DiNardo et al. (1996). The sample includes individuals aged 16 to 64. The straight lines in the kernel density plots are the median wages in both the non-emigrant and emigrant re-weighted sample.

(see Figure 3) is 1.90, which corresponds to a difference from non-emigrant average wages (1.96) of 0.068 ($SE = 0.001$), leaving an unexplained gap of 9.3 percentage points ($SE = 0.018$). For emigrants to Germany, the average wage in the re-weighted distribution is 1.90, meaning that the average observable selection is 0.066 ($SE = 0.008$) and the 1.9 percentage points ($SE = 0.009$) are due to unobservable selection. These differences indicate that whereas observable characteristics can explain 77 percent of the gap for emigrants to Germany, they can account for only 37 percent of that for emigrants to the UK.

3.4 Characteristics of unemployed non-emigrants versus emigrants

Table 4: Characteristics of unemployed individuals (non-emigrants and emigrants by destination)

	Non-Emigrants	Emigrants			
	(1)	All destinations (2)	UK (3)	Germany (4)	Other (5)
Education (years)	11.34	12.46	13.23	11.77	12.24
Primary(%)	19%	8%	5%	14%	8%
Secondary(%)	72%	71%	64%	74%	74%
Tertiary(%)	8%	20%	31%	11%	17%
Age	34	32	29	36	32
Females (%)	50%	30%	33%	28%	29%
Observations	206,064	733	221	185	327
Probability to emigrate		0.35%	0.1%	0.08%	0.15%

Source: Polish Labour Force Survey, years 1998 to 2008. Individuals who are unemployed, aged 16 to 64. Emigrants are Polish citizens who are now living in another country but who lived in Poland during the previous interviews of the survey. Primary educated are those who left education at age 16 (or less), secondary educated are those who left education between 17 and 20 years old and tertiary educated are those who left education any time after that age. Probabilities to emigrate are computed using only the sample of individuals who are unemployed. Entries are weighted with the survey weights for non-emigrants and the estimated weights for emigrants following Dustmann et al. (2015).

The dataset allows me to identify emigrants who were unemployed prior to their decision to emigrate. The unemployment rate in Poland ranged from 10 percent in 1998 to 19 percent before the EU Enlargement, down to 7 percent in 2008. It is plausible to expect that a part of the emigrant population was drawn from this sample. Table 4 shows the average characteristics of unemployed non-emigrants and emigrants by destination country. Selection on emigrant observable characteristics is in line with the findings shown in Table 1. In general, unemployed individuals in Poland have less education than individuals in employment; in fact, they are more likely to be primary educated. They are also younger and more likely to be female. In contrast, unemployed

emigrants are very similar to employed emigrants, yet with few differences in terms of education across destinations. Interestingly, unemployed individuals to the UK have more education (the share of tertiary is above 30 percent versus 21 percent in Table 1, column 3), while those to Germany are slightly less educated (the share of primary schooled is 14 percent among unemployed and 8 percent among employed, Table 1, column 4).

4 Explaining selection: the role of returns to skills

Following Borjas (1991), return to observable and unobservable skills has an influence on the selection of migrants, so in this section, I explore the differences in both observable and unobservable skill prices across source and destination countries. The price of education is usually measured as the wage return for each extra year of schooling.¹⁹ In the data, I approximate it for both Poland and the destination countries by the estimation of a simple Mincer regression. I thus estimate an OLS regression of (after tax) log hourly wages on years of education²⁰ using five age categories, marital status and gender, state/region of residence and year dummies for each country during the 1998-2008 period.²¹ I estimate this regression for all the workers and *additionally* also for only Polish immigrants in the two destination countries. I identify immigrants based on nationality designation in Germany and country of birth designation in the UK. I also consider two sub-samples of Polish immigrants: those who have recently arrived (i.e., have at least one year of residence) and the entire Polish immigrant population. For these latter, my OLS regression also controls for years since migration, to account for any quality differences across arrival cohort. As before, Table 5 reports the results for Poland (Panel A), the UK (Panel B) and Germany (Panel C), for the described samples.

The coefficients on years of education indicate that the market price of an additional year of school is higher in Poland (0.092) than in both the destination countries, although the return to education is slightly higher in the UK (0.058) than in Germany (0.054).²² Based on the implications

¹⁹Ibarraran and Lubotsky (2007), Fernández-Huertas Moraga (2013), Gould and Moav (2016) among the others.

²⁰Based on the suggestion by Heckman et al. (2003) that assuming linearity in education may not be appropriate, I also run the regressions using the three education categories described in previous sections; the ranking across countries remains the same.

²¹My age categories are 16-24, 25-34, 35-44, 45-54, and 55-64; the local divisions are 16 states in Germany, 20 regions in the UK and 16 regions in Poland.

²²The returns to education observed in this context are below those observed using gross wages because of taxation's tendency to compress the wage distribution, thereby lowering the return to education. The UKLFS provides data on (hourly) gross wages, which I use to re-estimate the returns to education. Here, the return is 20 percent higher (0.07), in line with the results reported in the literature (see e.g. Hogan and Rigobon 2002; Harmon and Walker 1995). Hence, assuming that the bias in the estimation of the return with net wage rather than gross is the same across countries, the results on selection would still hold. If the German system is more redistributive than the English system, then the returns to education in Germany using gross wages could be higher than in the UK. Were emigrants to be sorted based on gross wages (which is less likely, see above), the conclusions about selection might be different. Nevertheless, I reconstruct the gross wages for Poland and use another (comparable

Table 5: Mincer regressions

Dependent variable: Log hourly wages							
	A.Poland		B.UK		C.Germany		
	(1)	(2)	Polish immigrants	All	(5)	Polish immigrants	All
Recent			Recent			All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Education	0.092*** (0.000)	0.058*** (0.000)	0.024*** (0.007)	0.030*** (0.003)	0.054*** (0.000)	-0.018 (0.016)	0.028*** (0.003)
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.31	0.306	0.361	0.399	0.201	0.394	0.14
Root MSE	0.391	0.381	0.277	0.309	0.474	0.488	0.49
Observations	410,569	609,871	439	1,690	1,663,708	177	5,940

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. All samples includes individuals between the ages of 16-64 who are not self-employed. Equations are estimated from 1998 to 2008. Wages are real hourly net wages deflated using 2008 prices. Education is measured in years of education. Other controls include five age-category dummies (16/24 , 25/34, 35/44, 45/54, 55/64.), female dummy, marital status dummies (6 categories for Germany, 4 for the UK and 2 for Poland), region of residence dummies (16 regions for Germany, 20 regions for the UK and 16 regions for Poland), year dummies and years since migration when all immigrants are considered. Recent immigrants (those who arrived over the last year). Standard errors reported in brackets.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

of the model as reported in Borjas (1991), these estimates not only imply the negative selection of Polish emigrants to both destination countries, in terms of education, but also indicate a hypothetical positive sorting of emigrants to the UK with respect to Germany. This same pattern of negative selection from Poland to the two destination countries also holds when the return to education is estimated solely for Polish immigrants (columns 3, 4, 6 and 7). While, *prima facie*, returns to education are more likely to explain the average educational level of Polish emigrants to Germany, they do not seem informative about their observable selection to the UK. In fact, as reported in Table 1, emigrants to the UK are more similarly educated to non-emigrants, while emigrants to Germany have less education. On the other hand, in terms of sorting across country of destination, the results on the differential return to education imply slightly more negative sorting of emigrants to Germany relative to the UK, in line with the observed selection among the two groups of emigrants.

The Borjas (1991) model of migrant selection can be interpreted as a decomposition of the

dataset- GSOEP) for Germany and run the results on gross wages (see Section 5.1).

wage function not only into an observable part, defined by observable socio-economic variables but also into a part that results from unobservable characteristics. As a matter of fact, the model focuses primarily on how the selection is determined by the unobservable part of wages. The main challenge is to measure the return to unobservable skills. I follow Juhn et al. (1993) and Gould (2002) and more recently Gould and Moav (2016) on a similar topic, by using residual inequality to measure the price of unobservable skills. In Table 5, I report the residual variation measured by the Root MSE after controlling for the set of variables of the regression. It is interesting to note that the residual variation is higher in Germany (0.472 for all workers and 0.49 for Polish emigrants only) than in Poland (0.391) but lowest in the UK (0.38 and 0.30, respectively). According to the predictions of the theoretical model, individuals with above-average residual wages in the Polish population should emigrate to Germany, whereas those with below-average residuals wages should emigrate to the UK.²³

Following the same methodology as Gould and Moav (2016), I estimate residual wages using the following regressions, for a sample which includes both emigrants and non-emigrants, separately for each year (1998 to 2008) and occupation group (professional, service and blue collar) for a total of 33 regressions:²⁴

$$\log w_{ir} = \rho Education_{ir} + X'_{ir}\lambda + \theta_r + v_{ir} \quad (3)$$

where $\log w$ is the log hourly (net and real) wages; education is expressed in years of education; X is a vector of demographic characteristics, which include marital status, five age categories, female; and θ_r denotes regional fixed effects. Using these regressions, I estimate log-wage residuals (v_{ir}) as a measure of unobservable characteristics. Specifically, these wage residuals measure how distant each individual's (including each emigrant's) wage is from the average for that year and occupation, which allows me to empirically test whether the emigrant residuals are statistically smaller or larger than the non-emigrant residuals, both on average and by occupation. I find that the average difference in residual wages between non-emigrants and emigrants is positive and statistically significant for emigrants to the UK (negative selection) but negative and not significant for emigrants to Germany. By occupation, the differences in wage residuals for the UK are 0.237 ($SE = 0.087$) for professional workers, 0.228 ($SE = 0.048$) for service workers and 0.085 ($SE = 0.031$) for blue collar workers, all implying negative selection. For Germany, the difference

²³I have checked the results on the residual variance using the German Socio-Economic Panel (GSOEP) on a comparable sample using net wages: the results are confirmed, showing higher Root MSE for Germany with respect to the UK but also slightly higher returns to education (0.07). Dustmann et al. (2009) document an increase in residual wage inequality in Germany between 1975 and 2004, measured by the standard deviations of residuals, figures are in line with my findings.

²⁴I also estimate the residual wages running the same regression for each year separately, for a total of 11 regressions. Average residual wages for the UK emigrants and non-emigrants are the same, while average residual wages for German emigrants are slightly higher.

is positive and significant (negative selection) for professional and service workers at 0.350 ($SE = 0.134$) and 0.108 ($SE = 0.068$), respectively. In contrast, emigrants to Germany in blue collar jobs are more positively selected, with a wage residual difference of -0.063 ($SE = 0.031$). Hence, within occupational groups, the pattern of unobservable selection for Germany contains some ambiguity. When I focus solely on education, however, emigrants to the UK have more education than non-emigrants, in most cases. The differences in years of education between non-emigrants and emigrants are -1.15 ($SE = 0.45$) for professionals, -1.05 ($SE = 0.24$) for service workers and -1.02 ($SE = 0.14$) for blue collar workers. In Germany, however, only professional workers are significantly more positively selected on education than non-emigrants (-1.31, $SE = 0.69$), while both service workers and blue collar workers have, statistically, the same level of education as non-emigrants (with differences of 0.32 with $SE = 0.34$ and -0.12 with $SE = 0.14$, respectively). In Figure C.2, I report the residual distribution for the UK and Germany.

The probability to emigrate as a function of observable and unobservable skills shows clear differences across levels of education and below and above the average residual wages. Figure C.3 presents the share of emigrants by education and residual wages (top panels), and also within occupation categories (bottom panels). As revealed by the graphs, the rate of emigration to the UK increases with education, also within occupation groups, while it is negatively correlated with residual wages (it is higher for individuals whose residual wage is below the average). *Prima facie*, the rate of emigration to Germany is negatively correlated with education, but also with residual wages, with unclear patterns within occupations. I therefore test whether this relationship is statistically significant by running an OLS on the probability to emigrate, separately by destination country, where I control for gender, age, years of education, residual wages, year dummies (and regional dummies). Results confirm the positive but not significant relationship between the probability to emigrate and years of education, and the negative and significant relationship between probability to emigrate and residual wages for the UK. By contrast, the probability to emigrate is negatively and statistically significantly correlated with education while the positive correlation found in unobservable skills is not significant.

The relationships between education and the rate of emigration is therefore consistent with the predictions of the model only for Germany and not for the UK. On the contrary, the relationship between residual wages and the share of emigrants is consistent with the differences in unobservable prices for both destination countries, despite being not statistically significant for migrants to Germany. To test the model predictions, however, I also need variation in the returns to skills in the source versus the destination countries. Hence, similar to Gould and Moav (2016), I use the model described in equation 3 to estimate the returns to observable and unobservable skills within

the three occupational sectors and years in each country.²⁵ Table 6 reports the estimated returns to education, the standard deviation of residual wages, and the log real hourly wages expressed in zloty for each country and occupation separately, pooled across the years. The standard deviation of residuals is estimated by first computing the standard deviation by education within each year and occupation group. To do so, within each of the 33 groups, I take the standard deviation of each education group to account for a different educational composition within occupations and then estimate the average standard deviation by occupation and year. As Table 6 shows, the return to education across all occupations is much lower in both destination countries than in Poland and lower in the UK than in Germany. These lower returns should translate into negative emigrant selection in terms of education within occupation, and more negative for the UK. The pattern for relative returns to unobservable skills measured by the standard deviation of residuals, however, is more interesting: in all occupations, these returns are higher in Germany and lower in the UK than in Poland, raising an expectation that unobservable selection will be positive to Germany and negative to the UK.

Because the average probability of emigrating within occupations also depends on the wage gaps across countries, in Table 6, I also report the average log hourly real wages in zloty,²⁶ pooled for 1998 to 2008, to enable comparison across countries. The returns to migration (measured by the difference in wages between the destination countries and Poland) are high in all occupations because wages are higher in both destination countries than in Poland. Nevertheless, despite a very similar across-country occupational ranking, with professionals earning the highest wage in all three, blue collar workers earn the same wage as service workers in the UK, but higher wages in Poland and Germany. On the other hand, wage gaps are higher for professional and service workers in the UK, but higher for service and blue collar workers in Germany. These wage differentials could possibly explain the different scale effects, namely differences in the average emigration probability by occupation across the two destinations (see Table 1), which is higher than average for blue collar workers in Germany and for service workers in the UK.²⁷ One final observation is that gains also occur across occupations; that is, a blue collar worker in both destination countries earns much more than a professional in Poland. Hence, Polish emigrants would still profit from migration even in the event of occupational downgrading.²⁸

²⁵I am unable to exploit industrial sector variation because the industry information on employment pre-departure among emigrants is sporadic, so I would be missing many observations.

²⁶See Table C.8 for information on exchange rates across destination countries.

²⁷I use scatterplots to study the correlation between emigration rates and across-country wage differentials, by occupation (and education), I find a positive relationship for Germany, but not for the UK.

²⁸See Grogger and Hanson (2011) for more on the role of relative versus absolute differences in wages across countries. Using eq. 11 of the cited work, I compare the return to education and absolute wages in both destination and source countries. In the majority of cases, results would predict negative selection within occupations, with the exception of service workers for Germany.

Table 6: Average return to observable and unobservable skills by occupation (Poland, UK and Germany)

	Return to Education	Residual SD	log hourly wages (zloty)
	(1)	(2)	(3)
<i>Poland</i>			
Professionals	0.086	0.422	2.28
Service workers	0.055	0.397	1.79
Blue collars	0.039	0.367	1.84
<i>UK</i>			
Professionals	0.037	0.37	4.11
Service workers	0.020	0.31	3.62
Blue collars	0.011	0.31	3.62
<i>Germany</i>			
Professionals	0.049	0.525	3.76
Service workers	0.037	0.512	3.47
Blue collars	0.019	0.45	3.52

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Return to education and residual wages are estimated from a regression of log real hourly wages in zloty, controlling for 5 age-category dummies (15-25, 25-35, 36-45, 46-55, 56-64), gender, marital status (6 categories for Germany, 4 for the UK and 2 for Poland), region of residence (16 regions for Germany, 20 regions for the UK and 16 regions for Poland). In the immigrant regression, year since migration is added as a control. Regressions are estimated for years 1998 to 2008, for the all workers, aged 16 to 64, employed. Self-employed are excluded from all samples. Each regression is estimated separately for each occupation and year.

5 Estimating selection on observables and unobservables

To determine whether cross-country differences in returns to skills can explain the different reported relations between migration propensity and different skill levels, I use pre-migration emigrant occupation sector data to explain the probability that individual i who works in occupation j at time t decides to emigrate from Poland to country d . More specifically, any cost and scale effect on migration are approximated by various controls (years of education, age, sex, time and occupational dummies). Like Gould and Moav (2016), I begin by examining the link between emigration probability and first unobservable and then observable skills (defined as years of education) separately by destination country d . In doing so, I test whether selection within occupational sectors, in both unobservable and observable skills, is explicable by a systematic variation in the returns to skill prices across sectors within each country.

Selection on unobservables

For each destination d I run the following regression:

$$M_{ijt}^d = \gamma_0 + \gamma_1 x_{ijt} + \gamma_2 educ_{ijt} + \gamma_3 residual_{ijt} + \gamma_4 (SDpl - SDdest)_{jt} \times (residual)_{ijt} + \gamma_5 SDpl_{jt} + \gamma_6 SDdest_{jt} + \gamma_7 BETApl_{jt} + \gamma_8 BETAdest_{jt} + \tau_t + \alpha_j + \epsilon_{ijt}$$

where x_{ijt} is a vector of personal characteristics, including a gender dummy and five age categories, and $educ_{ijt}$ is the number of completed years of education for individual i . $residual_{ijt}$ is the individual's residual wage, calculated separately for each year and occupation using a standard Mincer wage regression (as described in Section 4). $SDpl_{jt}$ and $SDdest_{jt}$ are the standard deviations of Polish and destination country residuals computed from equation 3 (Section 4), estimated separately for each country. $BETApl_{jt}$ and $BETAdest_{jt}$ are the returns to education in occupation j and year t in Poland and in the destination country, respectively (calculated using the same regressions used for standard deviations of residuals). Finally, τ_t is a time dummy, α_j is a fixed-effect for occupational sector j and ϵ_{ijt} is the error term. The function is estimated using a linear probability model.²⁹

Residual wages within the sector are used to measure the unobservable skills, while the standard deviation is used to proxy returns to unobservable skills in sector j at time t . Years of education and returns to education proxy observable skills and returns to unobservable skills, respectively. The coefficient γ_2 measures the relationship between education and the probability of leaving the country, while the coefficient on residual wages (γ_3) measures the relationship between unobservable skills and emigration probability. It does so by indicating whether, within a sector, the emigration probability is increased or decreased by higher residuals. By explicitly controlling for education, and returns to education, when examining selection on unobservable skills, I separate the effect of unobservable skills from that of observable skills.

As the model shows, a lower (higher) return to skills in Poland relative to the destination country implies that higher (lower) skilled workers will leave the country, meaning that $\gamma_4 < 0$. For example, assuming that $SDpl_{jt} - SDdest_{jt}$ is negative, as in the case of Germany, individuals whose residual wages are above the average residual wages (i.e. above zero) should migrate. For those with residuals below the average, the migration probability should be smaller. Because I use occupational fixed effects to control for any heterogeneity in the migration probability across sectors and, by introducing year dummies, I control for differences in the emigration probability across years, the coefficient is identified by using variation in the difference in the standard deviations (between Poland and destination) within each occupation and year. The main identifying assumption is that (unobservable) heterogeneity in the emigration of higher skilled versus lower skilled workers is independent of the differences in returns to skills within each occupation.³⁰ The

²⁹The results are not affected by the model choice. Results estimated using a probit model are available upon request from the author.

³⁰This assumption implies that migration is not simultaneously affecting skill prices. Although using $t - 1$ skill

purpose is to test whether the emigration probability increases with skills in sectors (and years) in which relative returns to skills are higher in the destination country than in Poland and vice versa.

It is worth noting that occupational downgrading of migrants in the destination countries can occur because of several unobservable factors, including inefficiency of skill transferability or lack of complementary abilities (e.g. knowledge of the local language). This event can result into a downward bias of the returns to education, stemming from the negative correlation between observed characteristics and these unobservables. Because migrants may be aware of not securing a job in the same occupation once in the destination country, occupational downgrading may prevent returns to skills by occupation from being informative. In fact, if occupational downgrading is a severe issue and emigrants do not consider the differential levels of inequality across countries within the same sector because of that, then occupational downgrading should only weakens the effect of prices of skills (Gould and Moav 2016). I am not assuming that occupational downgrading is not happening, however in order to attenuate its effect on my results, I used a definition of occupation which is quite broad, where it is reasonable to assume that the switching costs of moving across broad occupations are high (given the investment in occupation-specific human capital). In addition, if occupational downgrading is temporary and most climb the occupational ladder after several years, what matters for long-term immigration is the prices of skills in the overall all population. Nevertheless, in the extensions, I test whether using the returns to skills of the Polish immigrant population in the destination country changes the results.

The top panel of Table 7 reports the results for selection on unobservable skills for the UK and Germany. Whereas columns 1 and 2 cover all the years analysed, columns 3 and 4 only reference the years from enlargement to just before the financial crisis (2004-2007). Because returns to skills are estimated using the regressions explained in the previous section, the standard errors are bootstrapped.³¹ The most interesting parameter is that for the interaction between the differences in returns to unobservable skills in the two countries and individual unobservable skills. The main finding is a negative coefficient for the UK that is driven by the time period, 2004-2007, a period that in fact yields the highest levels of emigration to the UK. For Germany, in contrast, the coefficients are positive and significant only when I consider the entire sampling period (column

prices (less affected by the migrant outflow) may partly solve this issue, results are the same but are not significant probably because of the loss of information, and therefore decrease the precision of the estimates.

³¹To test for non-linearities of sigma, following Gould and Moav (2016), I add a quadratic term for sigma, which turns out not to be significant. In another specification, I also control for occupational dummies linearly interacted with sigma, but the change in results is minimal; only a slight increase in the standard errors of the estimated coefficient of the difference. The coefficients of these interactions, however, are never significant, implying no differences across occupations in terms of sigma. This result is not surprising given that when constructing the prices of unobservable skills, I also consider any differences in standard deviation for each education group, as different education groups may have different levels of unobservables.

Table 7: Selection on unobservables and observable skills - UK and Germany

	OLS on probability to emigrate			
	1998-2008		2004-2007	
	(1)	(2)	(3)	(4)
	UK	Germany	UK	Germany
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.003*	0.005**	-0.014*	0.013
	(0.001)	(0.002)	(0.007)	(0.009)
B. Observable selection				
$(BET_{Apl} - BET_{Adest})_{jt} \times$ $(education)_{ijt}$	0.015***	0.003	0.017	0.005
	(0.003)	(0.002)	(0.010)	(0.005)
Observations	410,788	410,749	142,707	142,655
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, and residual wages, and only in Panel B regressions, occupation \times years of education dummies. The residual wages, residual SD and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

1), with no significant results reported in column 4. Interpreting these results in terms of the model predictions, I note that the sign of the coefficient for the UK is consistent with the model: lower returns to unobservable skills in the UK compared to Poland induce individuals with lower unobservable skills to leave Poland. Differently, the results do not follow the model predictions for Germany.

Selection on observables (education)

To test whether the probability of emigrating to the UK or to Germany is higher for highly educated individuals, when the returns to education in Poland are lower than those in the destination country i , I run the following OLS regression using the variables defined in the previous sub-section:

$$M_{ijt}^d = \gamma_0 + \gamma_1 x_{ijt} + \gamma_2 educ_{ijt} \times \alpha_j + \gamma_3 residual_{ijt} + \gamma_4 (BET_{Apl} - BET_{Adest})_{jt} \times (educ)_{ijt} + \gamma_5 SD_{pl_{jt}} + \gamma_6 SD_{dest_{jt}} + \gamma_7 BET_{Apl_{jt}} + \gamma_8 BET_{Adest_{jt}} + \tau_t + \epsilon_{ijt}$$

Because the level of education and differences in returns are higher in occupations where emigration probability is lower, controlling for the interaction between years of education and

occupational dummies allows me to account for a different educational composition within occupations that may mechanically drive the results. The coefficient of interest is thus γ_4 . The model predicts that lower returns to education in the destination country relative to Poland ($(BETA_{pl} - BETA_{dest})_{jt} > 0$) in sector j and year t will induce individuals with lower than average education in that sector and year to move. The parameter is identified using the variation across sectors and time of the differences in returns to education between countries, with controls for the distribution of education within occupations. The UK results (Table 7, Panel B) are positive and significant in column 1, indicating that higher returns to education in Poland than in the UK do not create an emigration incentive for individuals with below average education. In fact, they are more likely to do the opposite. This effect disappears, however, when I restrict the analysis to the post-enlargement period (column 3). The results from Germany, however, point to the fact that there are other factors driving migrant selection, as the results in Table 7 are not informative. Interestingly, no significant results on the role of return to education are found also in Gould and Moav (2016) but only on the occupation level analysis, like this case.

Interpretation

By more closely examining the magnitude of the coefficients obtained from the equation on unobservable selection, I can shed light on how the emigration probability changes with changes in the residuals. Hence, in Figures C.4 and C.5, I use the regression results (from Table 7, column 3) to estimate and plot the predicted post-enlargement migration probabilities to the UK (Figure C.4) and Germany (Figure C.5), for an individuals with average characteristics, to show how migration probability differs along the distribution of residual wages. More specifically, I show this graph for four different cases: (1) when the difference in returns is the average difference: 0.094 for the UK (Figure C.4A) versus -0.0815 for Germany (Figure C.5A); (2) when the difference in returns is equal to the median difference in the sample: 0.106 for the UK (Figure C.4B) but -0.074 for Germany (Figure C.5B); (3) when the difference in returns is equal to zero (Figures C.4C and C.5C); and (4) when the difference in returns is set at the average difference level of the other destination country: -0.0815 for the UK (Figure C.4D) compared to 0.094 for Germany (Figure C.5D). In these graphs, the probabilities are estimated at 9 points of the residual wage distribution (roughly distant 2.5 standard deviations from each other). For emigrants to the UK, the probability decreases with residuals when the difference in the returns to skills between Poland and the UK is positive (Figures C.4A and C.4B), is almost flat and never significant when the difference is zero, as there are no incentives to migrate (Figure C.4C) and increases with residuals when the difference is negative, as in the case of Germany (C.4D). This pattern is exactly what the model would predict: equalizing the residual variance across the two countries decreases the probability to zero, which implies that individuals find no benefit from migration, while decreasing inequality

in Poland with respect to the UK shifts the direction of unobservable selection towards the positive. Nevertheless, the shift in Figure C.4D is small and not significant, probably suggesting the need for a larger difference in the price of skills to induce positive emigrant selection. By increasing residual wages of emigrants to the UK to the same level of residual wages of non-emigrants (by 13 percentage points), for an average difference of skill prices (0.094), the probability to emigrate would decrease by 0.01 percentage point, around 30 percent of the baseline probability to emigrate to the UK as reported in Table 1. Yet Figure C.5 reveals a completely different pattern, in which the probability of emigrating to Germany is almost zero when the difference in returns is average or at the median, increases slightly (but never significantly) when returns in the two countries are the same (Figure C.5C) and only increases when the difference is at the UK's average.

5.1 Extensions

Returns to skills play a role in determining unobservable selection of emigrants; however, they are less informative on the selection on education for the UK, so not in line with the predictions of the Borjas (1987) model. In contrast, results for Germany are ambiguous. In order to uncover possible mechanisms, I firstly investigate the role of the reference group in the destination country, and define it by the Polish immigrants in the destination country. In fact, because of the occupational downgrading in both destination countries, the return to education of Polish immigrants is lower than the overall return to education in the labour force. As pointed out in the previous section, occupational downgrading creates a downward bias in the return to education because of the negative correlation between observed characteristics and unobservables, as evidenced in Table 5.³² I therefore use the return to unobservables and observables in the Polish immigrant population in the destination country as a proxy for the prices of skills.³³ The results reported in Table C.2 reveal some evidence on the role of return to education of Poles in the destination country, which I speculate could stem from pre-existing migration networks. In fact, networks can create more transparency to individuals in the source country with regard to the information on the return to education in the destination country: while the coefficients on residual inequality is statistically insignificant (columns 1 and 3), the coefficient on educational inequality (columns 2 and 4) is statistically significant with the expected negative sign supporting the predictions of the model. These results point to the fact that emigrants include inequality in their information set when making the decision to migrate, but they may change reference group depending on the

³²See for example Dustmann et al. (2013) and Rosso (2016) for the UK and Elsner and Zimmermann (2013) for evidence of EU8 immigrant downgrading in Germany.

³³It is worth noting that data limitations do not allow me to estimate this effect for the entire period of the analysis, as the number of Polish immigrants within occupations in the UK before 2004 is very low. Also, these estimates may suffer from low precision, because of the small samples of Polish immigrants in both the destination countries' surveys. In spite of this, results may still be informative of possible factors driving selection.

observability of the skill, selecting on observable returns of the Polish immigrant population, while considering the overall difference in residual inequality.

Moreover, one of the assumptions of the model is that the decision to emigrate to one country is independent from the decision to move to another. However, there may be some unobserved factors in the probability to emigrate which are correlated across destinations. For this reason, the reported results may be driven by the fact that the selection equation is estimated separately for each country.³⁴ In order to understand the role of other countries' inequality, I estimate the regressions including the differential return to skills of both destination countries, with the drawback that I am considering only two possible destinations (accounting for almost 50 percent of the total migration flows in the last years of the analysis). Results reported in Table C.3 are still informative about whether the baseline estimates reported in Table 7 are capturing any individual level heterogeneity correlated across destinations in the probability to emigrate. Interestingly, results are even stronger for unobserved selection in the UK for the period 2004-2007, while they remain positive yet not significant for Germany. In the case of observable selections, results are not conclusive, confirming the baseline.

Finally, the estimation of the return to skills within occupations may capture some endogenous behavior at the individual level and consequently some self-selection. Yet, there are reasons to believe that the occupational choice of migrants in the destination countries is often driven by the existing migration policies, as is the case in Germany, where a strong correlation between policies and migrant jobs is observed (see Section 2). In the sending country, on the other hand, emigrants endogenously select into occupations before they emigrate. For these reasons, at the cost of decreasing the variation in the return to skills, I estimate the return to skills and residual wages using equation 3 for each year separately (for a total of 11 regressions). Because of the decreased precision, results reported in Table C.4 are less likely to be significant, with the exception of the coefficient on residual inequality in the UK. In bottom part of the table, I estimate yearly return to skills in the population of Polish immigrant destination. This time, the coefficient on unobservable selection for the UK is more precisely estimated, and the coefficient for observable selection remains negative but significant only for the German case.

The literature on migrant selection shows that immigrants tend to sort mainly on post-tax wages and less on pre-tax. I confirm this finding by estimating the return to unobservable and observable skills, and the residual wages, with gross instead of net wages. Gross wages are not reported in the Polish LFS, but reconstructed from net wages, using the taxation rules and the information available on each household following the same procedure as in Dustmann et al. (2015). For Germany I use the GSOEP, a survey that contains information on gross wages, while

³⁴See Bertoli et al. (2013) on the role of destination country's earnings in the choice of location for Ecuadorian migrants.

gross wages are reported in the UKLFS. Results are reported in Table C.5. In general, the results are smaller in magnitude than the baseline results, but above all, they are not significant in any consistent way.

I further inspect the results by running a battery of checks on different samples. Return to skills are likely to matter more for individuals with higher labour market attachments and higher propensity to emigrate (for longer periods). In Table C.6 I restrict the sample to include: only individuals aged 25 to 55, individuals who have left for work-related reasons, males and eventually individuals who have been abroad for more than a year. As expected, results are mainly driven by younger individuals who move for working reasons. Results on males and on migrants abroad for more than a year are not precisely estimated, mainly because of the loss of information in selecting the sample, yet coefficients are reporting the same sign of the baseline.

Ultimately, the model predicts that selection is driven by the differential levels of inequality measured by the return to skill, while the scale effect (within occupations) is driven by the wage gaps across countries. In the estimated selection equations, occupational, education, year dummies and residual wages are already controlling for any scale effect, yet, I re-run the results including the average log wages by occupation and year (excluding residual wages) in Poland and in the destination countries. The signs of the estimated coefficients are confirmed, and are slightly lower in magnitude (Table C.7, top panel). Further, I check that unobserved heterogeneity across occupations and skills in the emigration probability across different labour markets defined by regions is not driving the results. Again, the coefficients confirm the previous results (Table C.7, middle panel). Lastly, I control for any time-varying productivity shocks across education groups by introducing the interaction between education and year dummies. Results for unobservable selection are confirmed, while those on observable selection are now significant also for the period 2004-2007 (Table C.7, bottom panel).

In the attempt to find some supporting evidence on the role of migration policies in defining occupational selection of migrants, I look at the distribution of occupations in the destination countries. Despite the differential levels of education of emigrants across destinations with Polish emigrants to the UK being more educated than the German ones, the distribution of Polish immigrants across occupations is very similar, with 16.17 percent of Poles working as professionals in the UK versus 21.6 percent in Germany, 20 percent versus 21.25 working as service workers and the remaining shares as blue collars. Moreover, the destination country surveys report similar distribution of education of Polish immigrants as the one found in Table 1 using the PLFS. I argue that this occupational selection in the destination countries can be driven by two different phenomena. Firstly, as reported in Section 2, more than 80 percent of Polish labour in Germany came through guest worker programmes, which had two requirements: language and educational.

Polish workers needed to have at least vocational training (secondary education) and speak some German. I relate these requirements with the some descriptive evidence from the data. As reported in Table 1, emigrants to Germany are overrepresented among secondary educated, with respect to other emigrants and non-emigrants; moreover, emigrants to Germany disproportionately came from *Opole* and *Lower Silesian*, regions that had higher than average shares of individuals learning German in school as a first language and highest share of biligual municipalities.³⁵ Secondly, in particular in the UK, Polish immigrants suffer from large occupational downgrading, as explained in the previous section. I hypothesize that this fact could possibly be rationalized by looking at blue collars' average wages. In both destination countries, an average blue collar worker would earn higher wages than an average professional worker in Poland (Table 6).

6 Conclusions

Taking advantage of the extremely high levels of Polish emigration to European countries between 1998 and 2008, I apply a simple framework of migrant selection (Borjas 1987; Borjas 1991) to a unique pre-migration labour market dataset in order to analyse the selection of emigrants with respect to the non-emigrant population. Not only does the dataset's inclusion of destination country allow me to link emigrant data with labour market data in each country, but I am also able to compare the wages of emigrants to UK and Germany with those of non-emigrants, and account for both observable and unobservable selection. I test for selection using both observable (education) and unobservable (wage residuals) skills and their prices. The price of the former is proxied by its returns and the latter is proxied by the wage residual standard deviation, both estimated using a simple wage regression.

While I report no education-based selection for Polish emigrants to the UK, they are more negatively selected on unobserved skills, consistent with the model's prediction as the wage residual standard deviation is higher in Poland than in the UK. My analysis thus documents a situation in which there is no selection on education despite negative selection on unobservable characteristics. As a matter of fact, observable characteristics can only explain a minor proportion of the wage variation for the UK, making unobservable selection more quantitatively relevant. In fact, when I estimate the role of differential returns to education between Poland and the UK, the sign of the estimated coefficients does not conform to the model predictions. Emigrants to Germany, in contrast, are more likely to fall in the middle of the distribution, but are not different from non-emigrants, despite residual inequality being lower in Poland than in Germany. An interesting preliminary result shows that, when the destination's reference group on which return to edu-

³⁵<https://rm.coe.int/language-education-policy-profile-poland-country-report/16807b3b4a>, accessed on 27 February 2019

cation are estimated is changed, and Polish immigrants are considered, observable inequality is more informative. I speculate that results may be related to the role of migrant networks in the destination country. Yet, this divergence may well be attributable to the different immigration policies in the two destination countries. Whereas most of the reported migration flow to the UK occurred after 2004, just after Poles were allowed to work in the UK without policy restrictions, most of the flow to Germany has been driven by the prevailing migration schemes. Indeed, several cases in the literature provide evidence that the Borjas (1987) model is more likely to hold when individuals are allowed to move across borders without restriction (Abramitzky et al. 2012; Gould and Moav 2016). The scale effect of migration also observed within occupations could therefore be related not only to the higher level of wages in the destination countries, but also to the programs in place at the time, attracting mainly workers with low occupational skills.

Yet, despite its weaknesses, my own application of the model to both observable and unobservable selection provides useful additional evidence on how migrants react to both labour market differences and differing migration policies across countries. Selection should be examined along different dimensions. My results show that when individuals are free to move across countries they are more likely to react to the economic incentives, as shown in the UK. This result is important as emigrant skills can affect the levels of inequality in the sending countries, but also impacts labour market integration in the destination country. At the same time, which reference groups inequality is measured on matters as well. Together with Borjas et al. (2019), this paper is one of the first to analyse emigrant selection in the European context, showing the importance of labour markets inequalities in determining the direction of selection. I analyse selection of emigrants to different destination countries, by using individual level labour market data before they emigrate, and relate them to the different labour market conditions and policies in the destination countries; however, further research in this direction should focus on understanding how these different factors interrelate with regard to migrant selection.

References

- Abramitzky, R., L. P. Boustan, and K. Eriksson (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review* 102(5), 1832–1856.
- Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *American Economic Review* 95(4), 960–980.
- Algan, Y., C. Dustmann, A. Glitz, and A. Manning (2010). The economic situation of first and second-generation immigrants in france, germany and the united kingdom. *The Economic Journal* 120(542), F4–F30.
- Ambrosini, J. W. and G. Peri (2012). The determinants and the selection of Mexico-US migrants. *The World Economy* 35(2), 111–151.
- Bertoli, S., V. Dequiedt, and Y. Zenou (2016). Can selective immigration policies reduce migrants quality? *Journal of Development Economics* 119, 100–109.
- Bertoli, S., J. Fernandez-Huertás Moraga, and F. Ortega (2013). Crossing the border: Self-selection, earnings and individual migration decisions. *Journal of Development Economics* 101(0), 75–91.
- Blanchflower, D. G. and C. Shadforth (2009). Fear, unemployment and migration. *The Economic Journal* 119(535), F136–F182.
- Borjas, G. (1991). National origin and the skills of immigrants in the postwar period. *NBER Working Paper No. 3575*.
- Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *The American Economic Review* 77(4), 531–553.
- Borjas, G. J., S. G. Bronars, and S. J. Trejo (1992). Self-selection and internal migration in the united states. *Journal of Urban Economics* 32(2), 159–185.
- Borjas, G. J., I. Kauppinen, and P. Poutvaara (2019). Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics. *The Economic Journal* 129(617), 143–171.
- Clark, K. and S. Drinkwater (2008). The labour market performance of recent migrants. *Oxford Review of Economic Policy* 24(3), 495–516.
- DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica* 64(5), 1001–1044.

- Drinkwater, S., J. Eade, and M. Garapich (2009). Poles apart? EU enlargement and the labour market outcomes of immigrants in the united kingdom. *International Migration* 47(1), 161–190.
- Dustmann, C., T. Frattini, and I. Preston (2013). The effect of immigration along the distribution of wages. *The Review of Economic Studies* 80(1), 145–173.
- Dustmann, C., T. Frattini, and A. Rosso (2015). The Effect of Emigration from Poland on Polish Wages. *The Scandinavian Journal of Economics* 117(2), 522–564.
- Dustmann, C., J. Ludsteck, and U. Schönberg (2009). Revisiting the german wage structure. *The Quarterly Journal of Economics* 124(2), 843–881.
- Duvell, F. (2004). Polish undocumented immigrants, regular high-skilled workers and entrepreneurs in the UK. A comparative analysis of migration patterns, strategies and identity process. *Prace Migracyjne, nr 54, Institute for Social Studies, Warsaw Universtiy*.
- Elsner, B. and K. F. Zimmermann (2013). 10 years after: EU enlargement, closed borders, and migration to germany. *IZA Discussion Paper No. 7130*.
- Fassmann, H. and R. Munz (1994). European east-west migration, 1945-1992. *International Migration Review* 28(3), 520–538.
- Fernández-Huertas Moraga, J. (2010). New evidence on emigrant selection. *Review of Economics and Statistics* 93(1), 72–96.
- Fernández-Huertas Moraga, J. (2013). Understanding different migrant selection patterns in rural and urban mexico. *Journal of Development Economics* 103(0), 182–201.
- Gould, E. D. (2002). Rising wage inequality, comparative advantage, and the growing importance of general skills in the united states. *Journal of Labor Economics* 20(1), 105–147.
- Gould, E. D. and O. Moav (2016). Does High Inequality Attract High Skilled Immigrants? *The Economic Journal*.
- Grogger, J. and G. H. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics* 95(1), 42–57.
- Harmon, C. and I. Walker (1995). Estimates of the economic return to schooling for the united kingdom. *American Economic Review* 85(5), 1278–1286.
- Heckman, J. J., L. J. Lochner, and P. E. Todd (2003). Fifty years of mincer earnings regressions. *NBER Working Paper No. 9732*.
- Hogan, V. and R. Rigobon (2002). Using heteroscedasticity to estimate the returns to education. *NBER Working Paper No. 9145*.

- Honekopp, E. (1997). Labour migration to germany from central and eastern europe. Old and new trends. *IAB Labour Market Research Topics*, 1–25.
- Ibarraran, P. and D. Lubotsky (2007). Mexican immigration and self-selection: New evidence from the 2000 mexican census. In *Mexican Immigration to the United States*, NBER Chapters, pp. 159–192. National Bureau of Economic Research, Inc.
- Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101(3), 410–442.
- Kaestner, R. and O. Malamud (2013). Self-selection and international migration: New evidence from mexico. *Review of Economics and Statistics*, forthcoming.
- Lee, D. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76(3), 1071–1102.
- Manacorda, M., A. Manning, and J. Wadsworth (2012). The impact of immigration on the structure of wages: Theory and evidence from britain. *Journal of the European Economic Association* 10(1), 120–151.
- Manning, A. and B. Petrongolo (2008). The part-time pay penalty for women in britain. *The Economic Journal* 118(526), F28–F51.
- Miera, F. (2008). Long term residents and commuters: Change of patterns in migration from poland to germany. *Journal of Immigrant & Refugee Studies* 6(3), 297–311.
- Parey, M., J. Ruhose, N. Netz, and F. Waldinger (2017). The Selection of High-Skilled Emigrants. *The Review of Economics and Statistics* 99(5), 776–792.
- Patt, A., J. Ruhose, S. Wiederhold, and M. Flores (2017). International Emigrant Selection on Occupational Skills. *CESifo Working Paper no. 6527*, 71.
- Rosso, A. (2016). Immigrant-native wage gap: the case of eastern european immigrants in the uk. *LdA Working Paper 405*,.
- Saleheen, J. and C. Shadforth (2006). The economic characteristics of immigrants and their impact on supply. *Bank of England Quarterly Bulletin Q4*, 375–385.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Trevena, P. (2009). 'New' polish migration to the UK: a synthesis of existing evidence. *ESRC Centre for Population Change Working Paper 3/2009*, 35.

A Sample construction

A.1 Polish Labour Force Survey

For the analysis, I use data from 1998 to 2008 and restrict the sample to those between 15 and 64 years old. I eliminate wage observations below the 1st percentile and above the 99th percentile to eliminate outliers. The variable on years of education is defined by re-coding the original variable in the survey (which was classified into nine categories): all individuals who have partially or fully completed primary school, or equivalently have 9 or less years of education (*“bez wykształcenia szkolnego,” “niepełne podstawowe”* and *“podstawowe”*; compulsory education in Poland starts at the age of 7 and ends at the end of 16, this is why I used 9 years of education as the first threshold); all those who have completed a vocational or general secondary education and have between 10 and 13 years of education (*“średnie zawodowe,” “średnie ogólnokształcące,” “gimnazjum”* and *“zasadnicze zawodowe”*); and all those who have a post-secondary or higher education, or more than 13 years of schooling (*“wyższe”* and *“policealne”*). Following this classification, I define primary educated as those with less than 9 years of education, secondary educated as those with between 10 and 13 years of education, and tertiary educated as those with more than 13 years. The survey reports net monthly wages and hours worked in the reference week. I compute an approximation of the hours per month by multiplying weekly hours by 4.3. Wages are expressed in 2008 prices, provided by the Central Statistical Office of Poland (GUS). Wages for emigrants are computed by average real wages using all the wage observation before the individual migrated. Occupational distribution of emigrant is recovered by using the occupational mode for each individual, again using all the available information before they left the country.

A.2 British Labour Force Survey

For the UK, I use the British Labour Force Survey. The sample period is 1998 to 2008. The sample is restricted to individuals aged 16-64 inclusive, men and women together; individuals who are still in school, are dropped from the sample. I define the immigrant population on the basis of the country of birth, as is usually done in the immigration literature (see Dustmann et al. 2013; Manacorda et al. 2012; Saleheen and Shadforth 2006 among others). The education variable is derived using age left full time education, assuming that individuals start school at the age of 5. The UKLFS definition of education in terms of the "age at which the individual left full-time education" has proven more reliable than any other education measure, especially for immigrants (Dustmann et al. 2013; Manacorda et al. 2012; Saleheen and Shadforth 2006). Earnings are computed from net weekly wages and divided by the number of hours worked in

the reference week, including overtime.³⁶ Wages are then transformed into real wages using 2008 CPI information provided by the Office for National Statistics (ONS). Real wages are trimmed below the 1st and above the 99th percentile to eliminate outliers. Wages are converted in zloty (from Table C.8) using information from the Polish National Bank. The ONS uses the Standard Occupational Classification (SOC) to classify occupations. There is a change in the occupational classification after 2001, from SOC1990 to SOC2000. I have recoded the occupation variable into the International Standard Classification of Occupations (ISCO-88) codes for both periods using tables provided by the Classification and Harmonisation Unit in the ONS, in order to be able to compare the distribution of occupations over years and across countries. Using the first digit of the code, I define three occupational sectors: professionals, service workers and blue collar (description in the text). For all tabulations and regressions, I use the sampling weights provided in the survey.

A.3 German Microcensus

The data source for Germany is the German Microcensus. For the analysis I use the Scientific Use Files for years 1998 to 2008 provided by the IZA Data Center. The sample period is 1998 to 2008. The sample is restricted to individuals aged 16-64 inclusive, men and women together, individuals who are still in school are dropped from the sample. The microcensus also reports net monthly income in brackets and it is right censored at 18,000 euro (35,000 Deutsche Mark before 2000). Immigrants are defined using their nationality, as no information on country of birth is available before 2005. The education variable is recoded in order to obtain the years of education. Following Algan et al. (2010) I assume that age left full time education is 16 for those without vocational training, 19 for those with vocational training and without a high school degree (*Abitur*), 22 for those with vocational training and a high school degree, and 25 for those with college education. In order to compare occupations across countries, I recode occupation in all the datasets following the International Standard Classification of Occupations (ISCO-88) codes. Using the first digit of the code, I define three occupational sectors: professionals, service workers and blue collar (description in the text). The variable on net monthly wages is not a continuous variable, wages are recorded in ranges and income is right censored at 18,000d (35,000 Deutsche Mark before 2000). The income is imputed as the midpoint of each income category following Algan et al. (2010). Wages are expressed at constant 2008 prices and transformed in zloty (from Table C.8) using the information on exchange rates from the Polish National Bank. For all tabulations and regressions sampling weights are used.

³⁶I have included the overtime hours following the Office for National Statistics, including them in the construction of the gross hourly wages reported in the dataset.

B Comparing the sample with full sample of emigrants

As pointed out before, the sample I consider for the wage analysis is not the *whole* sample available in the PLFS, which has been proved to be reliable in terms of ability to capture the relevant population of emigrants across destination countries (Dustmann et al. 2015). This smaller sample (hereafter *wage sample*) may not be fully representative of the emigrant population, as for example this group may be more likely to participate in the labour market. Given that this information is unobservable, I can still compare the characteristics of emigrants in both samples by destination (Table B.1). For this comparison, I use only the variables available in the full sample: education, gender and age. In Table B.2, I report results also for the regional distribution of emigrants in the two samples (where regions are ranked on the basis of the real average wages between 1998 and 2008, as explained in Section 3.3). Overall, the samples are almost statistically identical, and I reject the hypothesis of equal means only for few characteristics. Starting from the UK, there is a small but statistically significant difference in years of education, and the restricted sample is 0.54 years less educated than the full sample. This difference is likely to be driven by the lower share of tertiary educated. In terms of age, there is no statistically significant difference between the two groups. For emigrants to Germany, both samples are similarly educated, while they are more different with regard to age. The whole sample is almost two years younger than the wage sample, in particular the former group is twice as likely to be in the age group 16 to 25 (19 versus 9 percent), while in the latter group the age group 35 to 45 is more represented. I also check the gender distribution and, for both destinations, the share of females is lower in the wage sample. This is likely to be related to the gender differences in labour market participation rates and probabilities to be employed. In order to determine if there are differences in the regional distribution idiosyncratic to the wage sample that could affect the results, I have ranked regions according to the average log hourly wage between 1998 and 2008, then grouped them into four groups of four regions each, in the following order: regions below the bottom 25th percentile of the ranking, between the 25th and 50th percentile, between the 50th and the 75th percentile and above the 75th percentile. I look at the distribution of emigrants within these four region groups. Results are reported in Table B.2. I firstly consider emigrants to the UK. The wage sample of emigrants is overrepresented in the first group and it is underrepresented in the third group. For emigrants to Germany, the regional distribution shows a significant difference in the bottom regions and in the third groups where the wage sample is overrepresented, while they are underrepresented in second group.

So far, I have only described the differences in observable characteristics, without eliciting anything about how these differences are related to the wage distribution. For example, because of the lower share of tertiary educated in the restricted sample of emigrants to the UK, the reported

Table B.1: Comparing the restricted sample with the full sample of emigrants

	UK			Germany		
	All (1)	Wage sample (2)	Diff. (3)	All (4)	Wage sample (5)	Diff. (6)
Education (years)	13.267 (0.042)	12.721 (0.181)	0.546*** (0.186)	11.922 (0.032)	11.681 (0.171)	0.241 (0.174)
Primary	0.036 (0.003)	0.064 (0.017)	-0.028 (0.018)	0.080 (0.004)	0.094 (0.022)	-0.014 (0.022)
Secondary	0.676 (0.008)	0.725 (0.033)	-0.049 (0.034)	0.808 (0.006)	0.820 (0.032)	-0.012 (0.032)
Tertiary	0.287 (0.008)	0.211 (0.030)	0.077** (0.031)	0.112 (0.005)	0.086 (0.025)	0.027 (0.026)
Female (%)	0.391 (0.008)	0.287 (0.033)	0.104*** (0.034)	0.329 (0.007)	0.189 (0.033)	0.140*** (0.034)
Age (years)	28.826 (0.131)	29.194 (0.544)	-0.368 (0.559)	34.126 (0.157)	36.045 (0.788)	-1.919** (0.803)
Age 16-25	0.307 (0.008)	0.271 (0.032)	0.035 (0.033)	0.193 (0.006)	0.097 (0.022)	0.096*** (0.023)
Age 25-35	0.520 (0.009)	0.550 (0.036)	-0.031 (0.037)	0.381 (0.008)	0.384 (0.040)	-0.003 (0.041)
Age 35-45	0.103 (0.005)	0.106 (0.023)	-0.003 (0.024)	0.228 (0.006)	0.296 (0.037)	-0.068* (0.037)
Age 45-55	0.059 (0.004)	0.072 (0.019)	-0.013 (0.020)	0.163 (0.006)	0.196 (0.031)	-0.033 (0.032)
Age 55-64	0.011 (0.002)	0.000 (0.000)	0.011*** (0.002)	0.034 (0.003)	0.027 (0.013)	0.007 (0.013)

Source: Source: Polish LFS, years 1998 to 2008. Individuals aged 16 to 64. All refers to the whole sample of emigrants, wage sample is the sample of emigrants for whom I observe positive wages. Primary educated are those who left education at age 16 (or less), secondary educated are those who left education between 17 and 20 years old and tertiary educated are those who left education any time after that age. Standard errors reported in brackets, clustered by region.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level and they are reported only for the difference columns as all the mean values (columns 1,2,4,5) are significant at 1%.

average wage is likely to have a downward bias compared to the wage I would have observed if the wage distribution for the full sample of emigrants had been observed (assuming there are no

Table B.2: Comparing the all emigrant sample with the wage sample: regional distribution

	UK			Germany		
	All (1)	Wage sample (2)	Diff. (3)	All (4)	Wage sample (5)	Diff. (6)
below 25th	0.28	0.40	-0.125** (0.032)	0.23	0.34	-0.112*** (0.033)
25th to 50th	0.36	0.33	0.026 (0.034)	0.48	0.32	0.168*** (0.039)
50th to 75th	0.16	0.10	0.06** (0.02)	0.14	0.21	-0.076** (0.068)
above 75th	0.19	0.16	0.037 (0.028)	0.144	0.12	0.020 (0.026)

Source: Source: Polish LFS, years 1998 to 2008. Individuals aged 16 to 64. Full sample is the whole sample of emigrants, restricted sample is the sample of emigrants for whom I observe positive wages. Regions are ranked according to the distribution of regional average log hourly wages (in 2008 prices), between 1998 and 2008. Each group includes the following regions, below the 25th percentile: Kuyavian-Pomeranian, Swietokrzyskie, Warmian-Masurian, Łódź and Greater Poland; between the 25th and 50th percentile: Subcarpathian, Lubusz, Lublin, West Pomeranian, Pomeranian, Opole; between the 51th and 75th percentile: Podlaskie and Lower Silesian; above 76th percentile: Lesser Poland, Silesian and Masovian. Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

differences in unobserved characteristics). In this case, the wage-gap between emigrants to the UK and non-emigrants is overestimated. Similarly, the lower share of females in the restricted sample of both destination countries may create an upward bias in the observed wage and the actual wage of emigrants before emigration would be even lower. In this case, the wage gap between emigrants and non-emigrants is underestimated. How are these differences translated into the wage distribution? In order to assess that, I perform a re-weighting exercise on the wage sample so that it matches the characteristics of the sample with all the emigrants. I use the nonparametric reweighting procedure introduced by DiNardo et al. (1996), where the distribution of wages of emigrants in the wage sample is reweighted so that the frequencies of the characteristics (age, education, gender and region of previous residence) resemble those of the whole sample. The resulting distribution is the one I would have observed if the information on wages was available for the whole sample of emigrants. The main assumption for this exercise is that the reason why wages for all emigrants are observed is random and it is not related to any unobservable factor (i.e. labour market attachment). The distributions of log hourly wages are estimated using a kernel. More formally, I decompose the density of log hourly wages w into a price $g(\cdot)$ and a composition $h(\cdot)$ functions. The price function simply represents how a change in observable characteristics, denoted by x , is translated into wages, whilst the composition function is the density of characteristics. I know the density function for both groups: wage sample (W) and the

whole sample (all).

The wage density for individuals in sample i in Poland (PL) is given by

$$f_{i,PL}(w) = \int g_{i,PL}(w|x)h_{i,PL}(x)dx; i = W, all \quad (\text{B.1})$$

Yet, I only observe wages for the restricted sample, as I do not observe $f_{all,PL}(w)$. This can be estimated using the density of characteristics which is known and the price function of the other group.

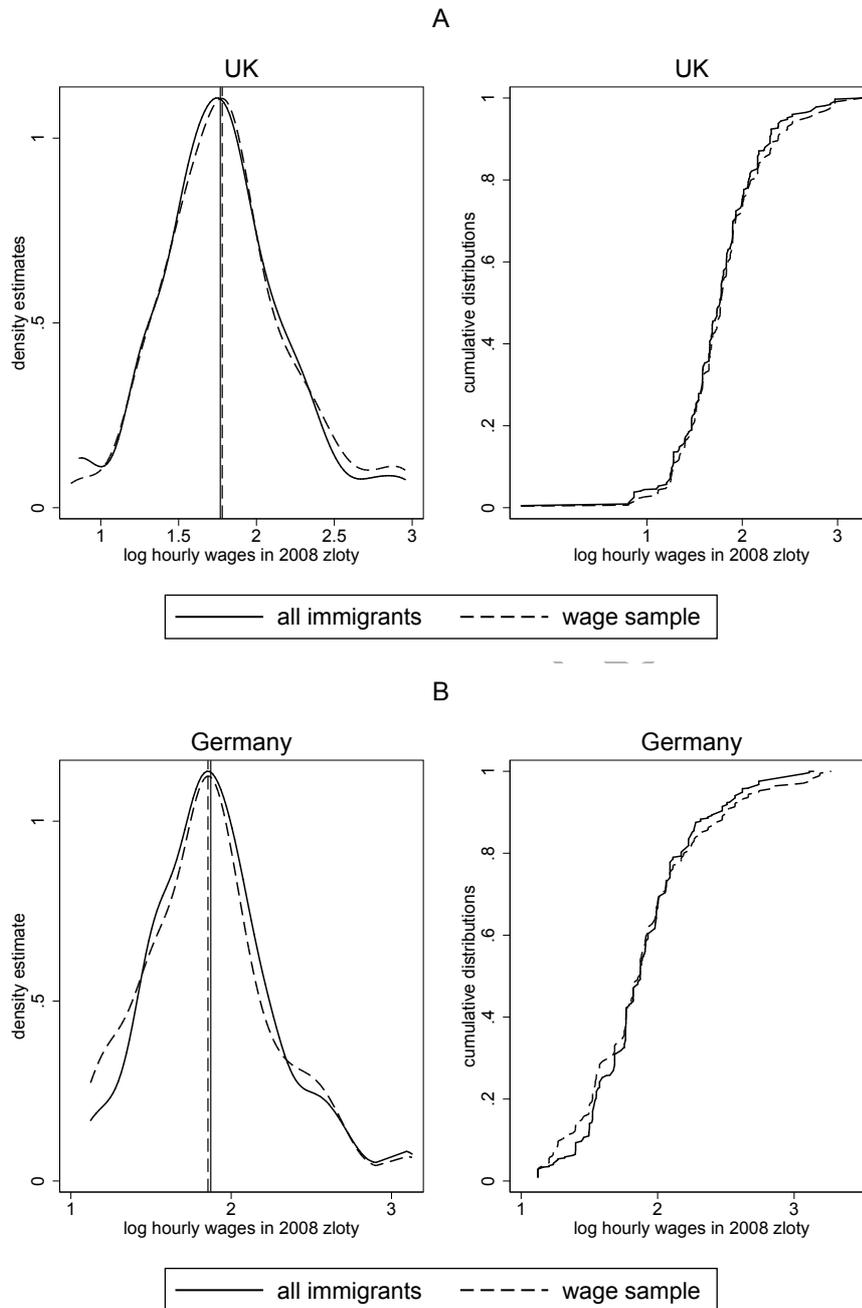
$$\hat{f}_{all,PL} = \int g_{W,PL}(w|x)h_{all,PL}(x)dx. \quad (\text{B.2})$$

DiNardo et al. (1996) suggest reweighting the density of the restricted sample by the ratio $\theta = \frac{h_{all,PL}(x)}{h_{W,PL}(x)}$, so that the density is estimated by

$$\hat{f}_{all,PL} = \int g_{W,PL}(w|x)h_{W,PL}(x)\frac{h_{all,PL}(x)}{h_{W,PL}(x)}dx \quad (\text{B.3})$$

Following Bayes' rule, the ratio can be calculated and estimated: $\theta = \frac{h_{all,PL}(x)}{h_{W,PL}(x)} = \frac{\frac{Pr(all|X)}{1-Pr(all|X)}}{\frac{Pr(all)}{1-Pr(all)}}$. I estimate $Pr(all|X)$ using a probit model of the probability of being in the full sample regressed on the observable characteristics (years of education, five age categories, gender dummy, regional and year dummies), whilst $Pr(all)$ is just the probability of being in the whole sample. I estimate the log hourly real wage distribution for emigrants to both the destination countries. The results are reported in Figure B.1.

Figure B.1: Kernel density estimate of log hourly real wages and differences



Source: Polish Labour Force Survey, years 1998 to 2008. For the estimation of the kernel densities on the left-hand side plots, I use the Gaussian kernel function as in DiNardo et al. (1996). I use the optimal bandwidth rule from Silverman (1986). The right-hand side plots report the difference between the density estimates ($f_{F,PL}(w) - \hat{f}_{R,PL}$). The sample includes individuals aged 16 to 64. "all immigrants" is the wage distribution of the wage sample re-weighted by the whole sample characteristics, "wage sample" is the distribution of the sample of emigrants with positive wage information. The straight lines in the kernel density plots are the median wages in both the full and restricted sample.

The plots on the right of both A and B report the wage distribution, while the panels on the left report the difference. At the top, I show the results for emigrants to the UK (A) while at the bottom for emigrants to Germany (B). The straight lines in the plots on the right of both A and B are the median values of the wage distribution in the wage and re-weighted sample. There

are no striking differences for Polish emigrants to the UK. In particular, if all emigrants were paid according to the same prices as the wage sample, their wage distribution would be slightly more dense below the median and just above the median, certainly less dense at the top, with no differences at the bottom. This shows that even if all emigrants are on average more educated the wage distribution would hardly be affected (both median and mean wages are lower in the reweighted sample), which is reassuring as the risk of overestimating the emigrant/non-emigrant wage gap is low. Panel B of Figure B.1 reports the results from the same exercise, but for Poles in Germany. In this case, there is some negative mass at the bottom of the distribution, where the wage sample is more represented, positive around the median and with almost no difference at the top of the distribution (this is also reflected in the higher median (2 log points) and much higher mean wages (4 log points) in the re-weighted distribution).

C Extensions

C.1 Estimating bounds of unobservables

In this section I provide evidence of the effect that the emigration of single households may have on unobservable selection, on average, by trimming the residual wages at the bottom and at the top of the distribution, which allows me to set bounds for the average residual wage in the presence of sample selection, like in the case of missing single households who decided to emigrate. In order to compute the bounds, I need to assume that, for each year and region in Poland, the proportion of non-emigrant individuals living in single households is the same as the proportion found among all emigrants. The average share of individuals living in single households in Poland between 1998 and 2008 is 0.06. In a similar fashion to Lee (2009), the proportion of nonmissing observations among all emigrants is therefore 0.94 (1-0.06) on average, while I assume it to be 1 among non-emigrants (there are no missing observations among the non-emigrants). This implies that the trimming proportion is 0.06 in the emigrant group³⁷. For each year and region, I compute the p th quantile: the upper bound for the emigrant residual wages is therefore the mean after trimming the tail of the distribution below the p th quantile. After trimming, the result mean for emigrants to Germany is 0.048 (almost 10 times the average) but remains negative and lower in magnitude (7 times smaller) for the UK. Symmetrically, I trim the distribution above the $1-p$ th quantile and compute the average residual to obtain the lower bounds: the lower bound for emigrants to Germany is negative in this scenario, while for emigrants to the UK it is larger in magnitude than the average residual wage and therefore more negative. This exercise is informative on the bounds we should expect in the selection of emigrants under two scenarios if I observed single

³⁷ $p = \frac{1-0.96}{1}$

households who emigrated: the first one in which all the households did not belong to the upper part of the residual wage distribution (lower bounds) and the second in which all the households did not come from the lower part of the residual wage distribution (upper bounds). This would still imply negative selection on unobservable factors for emigrants to the UK, while the results are ambiguous for emigrants to Germany.

Table C.1: Upper and lower bounds of unobservables

	Average	Upper bound	Lower Bound
UK	-.136	-.018	-.186
Germany	.005	.048	-.082
Other	-.047	.003	-.107

Source: Polish LFS, Note: Residual wages, residual SD and return to education are estimated using a regression of log hourly *net* wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Upper bounds are estimated assuming that the average of percentage of single households (p) is missing below the p th percentile distribution, while lower bounds assume that the $100 - p$ th percentile of the distribution (represented by the share) is missing in the distribution.

C.2 Tables on Extensions and Robustness

Table C.2: Selection on unobservables and observables, for UK and Germany (skill prices of Polish immigrants)

	OLS on probability to emigrate, 2004-2007			
	UK		Germany	
	Unobservable	Observable	Unobservable	Observable
	(1)	(2)	(3)	(4)
$(SD_{pl} - SD_{dest})_{jt} \times (residual)_{ijt}$	-0.0002 (0.0001)		0.00005 (0.0006)	
$(BETA_{pl} - BETA_{dest})_{jt} \times (education)_{ijt}$		-0.001** (0.0004)		-0.0009* (0.0005)
Observations	142,707		142,655	
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, residual wages, and only for observable selection, occupation \times years of education dummies. Residual wages, residual SD, and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table C.3: Selection on unobservables and observable skills - UK and Germany

	OLS on probability to emigrate	
	1998-2008 (1)	2004-2007 (2)
A. Unobservable selection		
$(SD_{pl} - SD_{uk})_{jt} \times$ $(residual)_{ijt}$	-0.003 (0.004)	-0.027* (0.016)
$(SD_{pl} - SD_{ger})_{jt} \times$ $(residual)_{ijt}$	0.002 (0.004)	0.032 (0.019)
B. Observable selection		
$(BET_{apl} - BET_{auk})_{jt} \times$ $(education)_{ijt}$	0.016*** (0.003)	-0.035 (0.038)
$(BET_{apl} - BET_{ager})_{jt} \times$ $(education)_{ijt}$	-0.0002 (0.005)	0.034 (0.029)
Observations	410,950	142,808
Occupation FE	Yes	Yes
Controls	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, and residual wages, and only in Panel B regressions, occupation \times years of education dummies. The residual wages, residual SD and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table C.4: Selection on unobservables and observable skills, yearly variation - UK and Germany

	OLS on probability to emigrate			
	1998-2008		2004-2007	
	(1)	(2)	(3)	(4)
	UK	Germany	UK	Germany
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.003** (0.001)	-0.003 (0.005)	0.051* (0.03)	0.014 (0.039)
B. Observable selection				
$(BET_{Apl} - BET_{Adest})_{jt} \times$ $(education)_{ijt}$	0.0004 (0.001)	0.003 (0.003)	0.012 (0.014)	0.010 (0.008)
Skill prices of Polish immigrants in destination				
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.001*** (0.0003)	-0.0005 (0.001)	-0.027** (0.010)	-0.004 (0.003)
B. Observable selection				
$(BET_{Apl} - BET_{Adest})_{jt} \times$ $(education)_{ijt}$	0.0005 (0.0006)	-0.002*** (0.0007)	0.0009 (0.001)	0.020 (0.015)
Observations	410,788	410,749	142,707	142,655
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, and residual wages, and only in Panel B regressions, occupation \times years of education dummies. The residual wages, residual SD and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table C.5: Selection on skills using gross wages- UK and Germany

	OLS on probability to emigrate			
	1998-2008		2004-2007	
	(1)	(2)	(3)	(4)
	UK	Germany	UK	Germany
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.001 (0.001)	0.001 (0.001)	-0.002 (0.002)	0.011** (0.004)
B. Observable selection				
$(BET_{Apl} - BET_{Adest})_{jt} \times$ $(education)_{ijt}$	0.011*** (0.002)	-0.0008 (0.002)	0.015 (0.013)	-0.008*** (0.005)
Observations	410,788	410,739	142,707	142,654
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, GSOEP, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, residual wages, and only for observable selection occupation \times years of education dummies. Residual wages, residual SD and return to education are estimated using a regression of log hourly *gross* wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets.* indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table C.6: Selection on unobservables and observable skills - UK and Germany

	OLS on probability to emigrate			
	1998-2008		2004-2007	
	(1)	(2)	(3)	(4)
	UK	Germany	UK	Germany
Age group 25-55				
A. Unobservable selection				
$(SDpl - SDdest)_{jt} \times$ $(residual)_{ijt}$	-0.005*** (0.001)	0.006** (0.002)	-0.020*** (0.007)	0.018 (0.013)
B. Observable selection				
$(BETApl - BETAdest)_{jt} \times$ $(education)_{ijt}$	0.016*** (0.003)	0.002 (0.003)	0.008 (0.012)	0.0003 (0.006)
Observations	367,846	367,848	128,630	128,610
Labour migrants				
A. Unobservable selection				
$(SDpl - SDdest)_{jt} \times$ $(residual)_{ijt}$	-0.003* (0.001)	0.005** (0.002)	-0.014* (0.008)	0.012 (0.011)
B. Observable selection				
$(BETApl - BETAdest)_{jt} \times$ $(education)_{ijt}$	0.014*** (0.003)	0.002* (0.001)	0.014 (0.010)	0.007 (0.005)
Observations	410,783	410,746	142,705	142,653
Males				
A. Unobservable selection				
$(SDpl - SDdest)_{jt} \times$ $(residual)_{ijt}$	-0.004* (0.002)	0.017 (0.003)	-0.008 (0.015)	-0.004 (0.003)
B. Observable selection				
$(BETApl - BETAdest)_{jt} \times$ $(education)_{ijt}$	0.021*** (0.005)	0.012**** (0.004)	0.052** (0.021)	0.023** (0.009)
Observations	219,449	219,438	75,929	75,904
Abroad for more than one year				
A. Unobservable selection				
$(SDpl - SDdest)_{jt} \times$ $(residual)_{ijt}$	-0.0005 (0.0005)	0.004*** (0.001)	-0.002 (0.002)	0.011*** (0.004)
B. Observable selection				
$(BETApl - BETAdest)_{jt} \times$ $(education)_{ijt}$	0.002* (0.001)	-0.001* (0.0006)	0.007 (0.003)	0.00042 (0.001)
Observations	410,611	410,619	142,568	142,564
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, and residual wages, and only in Panel B regressions, occupation \times years of education dummies. The residual wages, residual SD and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table C.7: Selection on unobservables and observable skills - UK and Germany

	OLS on probability to emigrate			
	1998-2008		2004-2007	
	(1)	(2)	(3)	(4)
	UK	Germany	UK	Germany
Controlling for average wages				
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.005*** (0.001)	0.0005 (0.0007)	-0.008*** (0.003)	0.003** (0.001)
B. Observable selection				
$(BETA_{pl} - BETA_{dest})_{jt} \times$ $(education)_{ijt}$	0.006*** (0.001)	0.002 (0.001)	0.004 (0.004)	0.0007 (0.002)
Observations	410,788	410,7498	142,707	142,655
Controlling regional dummies				
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.003* (0.001)	0.005*** (0.002)	-0.014* (0.008)	0.013*** (0.010)
B. Observable selection				
$(BETA_{pl} - BETA_{dest})_{jt} \times$ $(education)_{ijt}$	0.015*** (0.003)	0.003 (0.002)	0.017* (0.010)	0.005 (0.005)
Observations	410,788	410,7498	142,707	142,655
Controlling year dummies \times years of education				
A. Unobservable selection				
$(SD_{pl} - SD_{dest})_{jt} \times$ $(residual)_{ijt}$	-0.003 (0.002)	0.005** (0.002)	-0.014* (0.009)	0.013 (0.01)
B. Observable selection				
$(BETA_{pl} - BETA_{dest})_{jt} \times$ $(education)_{ijt}$	0.005*** (0.001)	.001 (0.001)	0.023** (0.01)	0.008** (0.002)
Observations	410,788	410,749	142,707	142,655
Occupation FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Source: Polish LFS, UK LFS, German Microcensus, 1998 to 2008. Individuals aged 16-64. Other controls include: 5 age-category dummies, gender dummy, year dummies, return to education in Poland and in the destination country, residual standard deviation in Poland and in the destination country, years of education, and residual wages, and only in Panel B regressions, occupation \times years of education dummies. Results at the bottom of the table also include year dummies \times years of education. The residual wages, residual SD and return to education are estimated using a regression of log hourly wages on years of education, five age categories, gender and region of residence fit, separately, for each year (1998 to 2008) and occupation. Bootstrapped standard errors reported in brackets. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

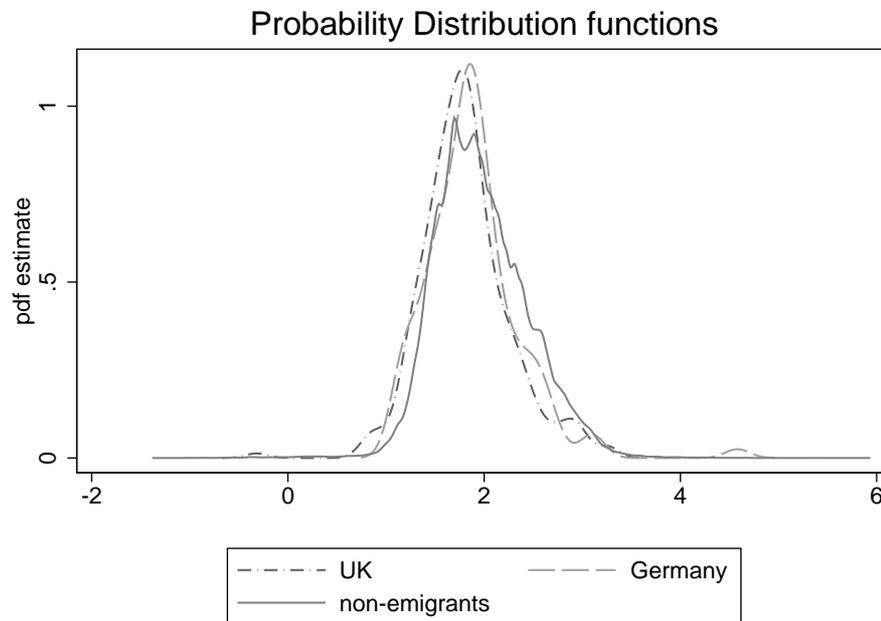
Table C.8: Zloty Exchange
Rates to € and £

	PNL\€	PNL\£
1998	3.9	5.76
1999	4.23	6.42
2000	4	6.58
2001	3.66	5.89
2002	3.84	6.11
2003	4.39	6.35
2004	4.54	6.7
2005	4.02	5.88
2006	3.89	5.71
2007	3.79	5.54
2008	3.51	4.43

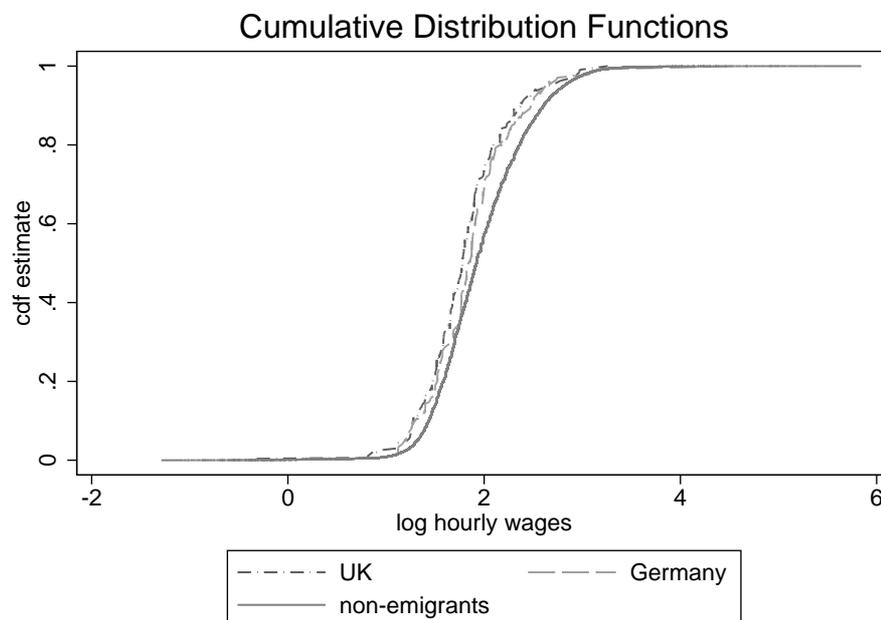
Source: Polish National Bank, statistics on exchange rates (Exchange rates archive).

Figure C.1: PDF and CDF estimates of log hourly real wages.

A

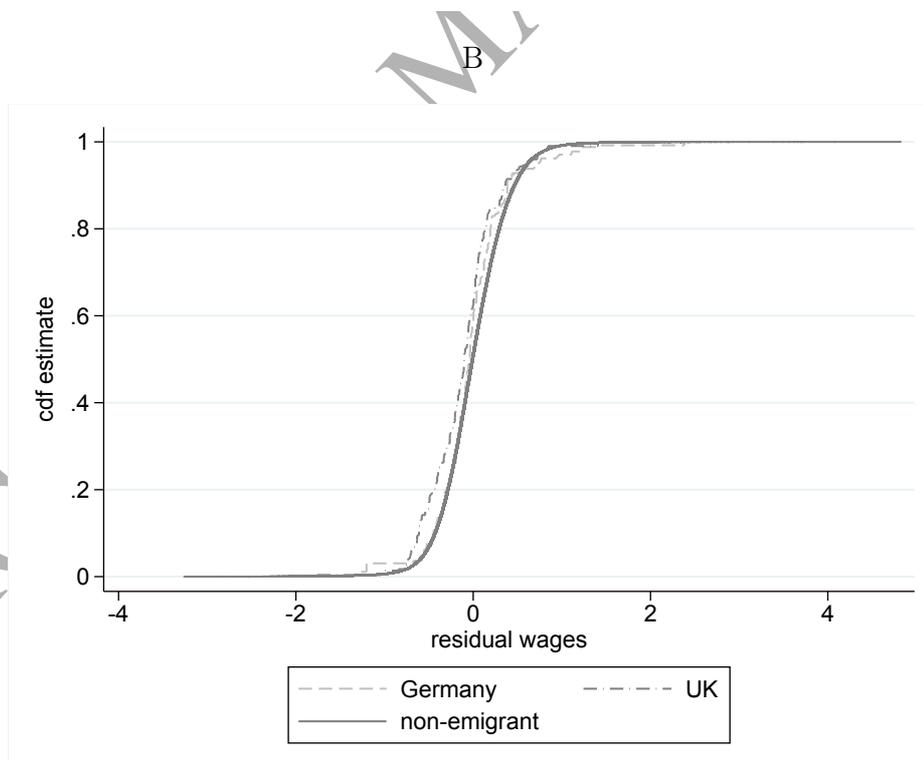
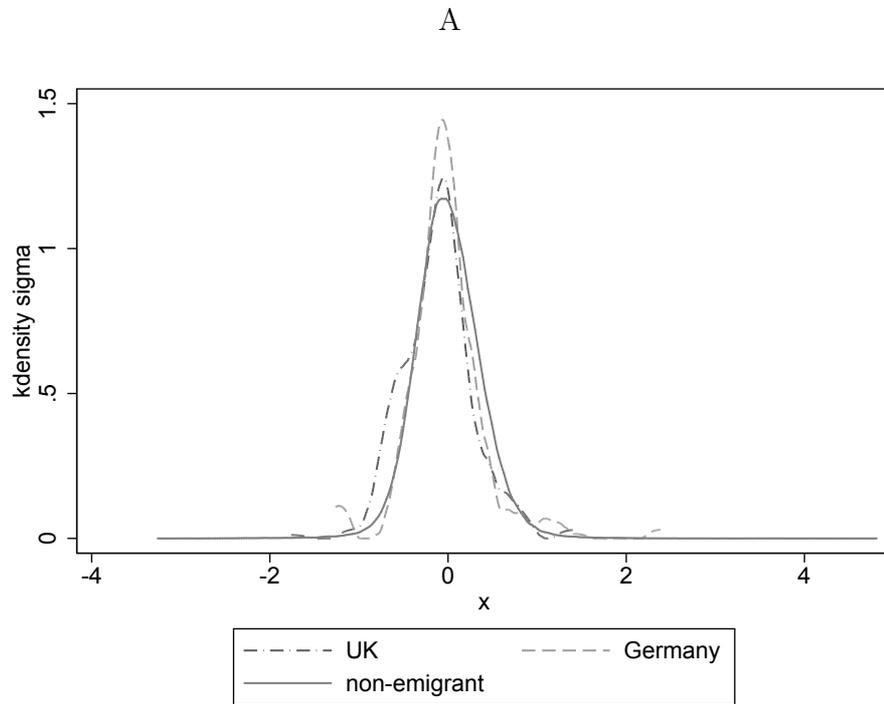


B



Source: Polish Labour Force Survey. Log hourly wages (in 2008 prices) pooled over the 1998-2008 period for the non-emigrant population (solid line), emigrants to Germany (dashed line) and emigrants to the UK (dashed-dotted line). For the estimation of the kernel densities on the top plot, I use the optimal bandwidth rule introduced by Silverman 1986.

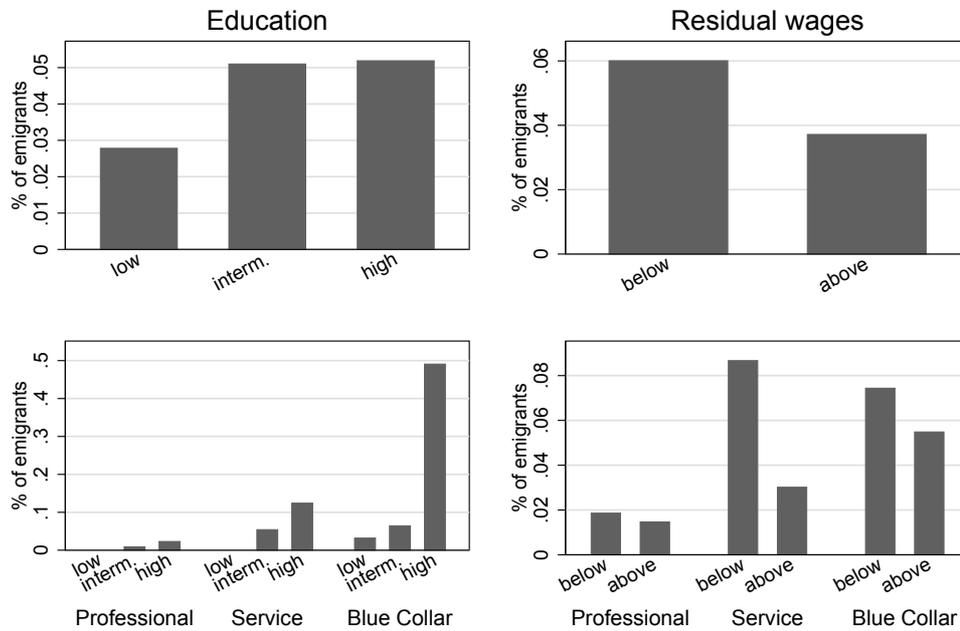
Figure C.2: PDF and CDF estimates of residual wages



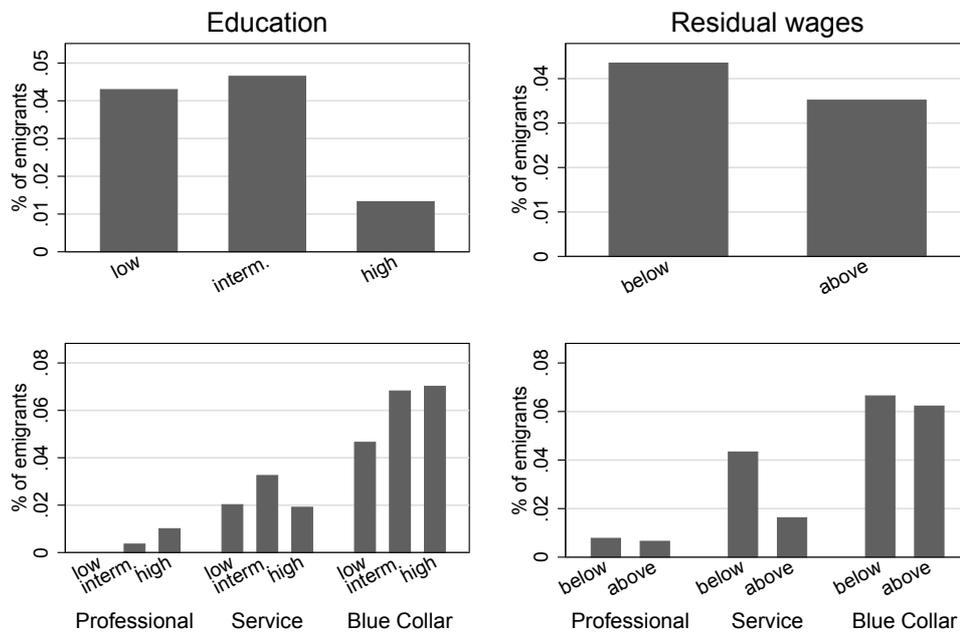
Source: Polish Labour Force Survey. Residual wages (estimated as reported in Section 4, equation 3) for the non-emigrant population (solid line), emigrants to Germany (dashed line) and emigrants to the UK (dashed-dotted line). For the estimation of the kernel densities on the top plot, I use the Gaussian kernel function with optimal bandwidth rule introduced by Silverman (1986).

Figure C.3: Probabilities of emigration by wage residuals and education.

UK

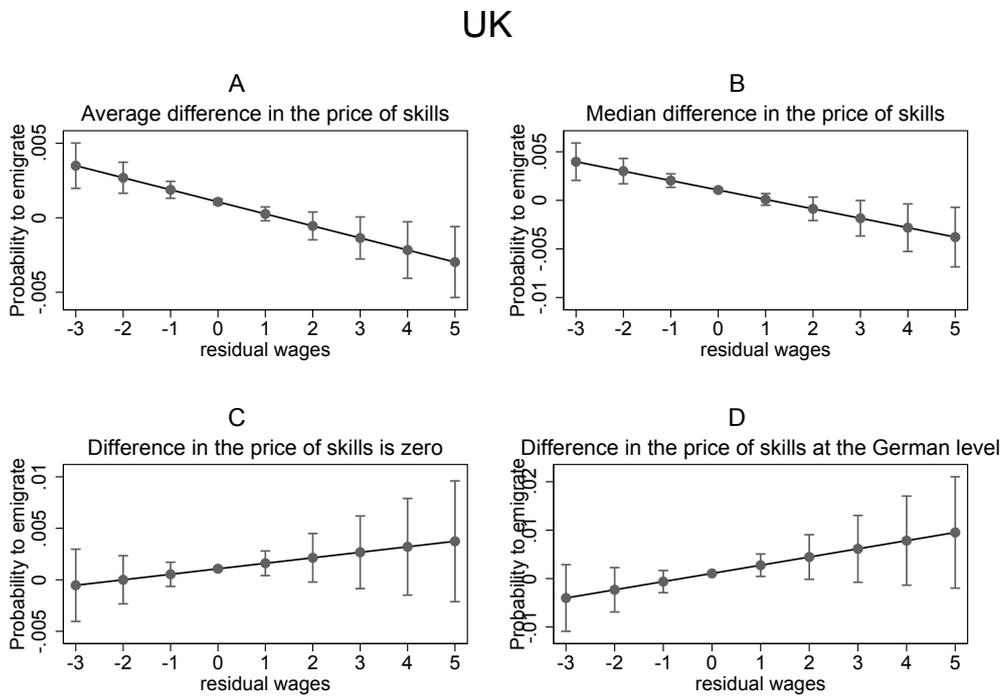


Germany



Source: Polish Labour Force Survey. Below and above refers to below or above the average residual wage (zero)

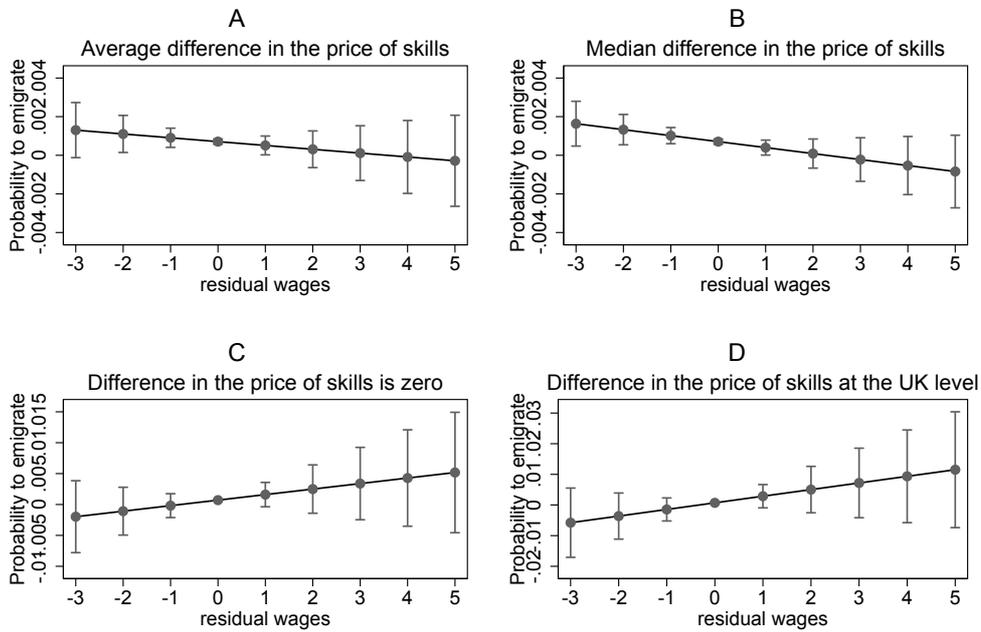
Figure C.4: Predicted probabilities of emigration by wage residuals and different levels of differences in the standard deviations.



Source: OLS estimates from Table 7, column 1 and Polish LFS for the estimation of residual wages using a regression of log hourly wages on years of education, five age categories, marital status, gender and region of residence estimated for each year from 2004 to 2007. The difference in returns is either the average difference between the residual variation in Poland and in the destination country in the sample (Panel A: 0.094), the median difference in the sample (Panel B: 0.106), the difference is zero (Panel C) or is at the same average difference level of the other destination country (Panel D: -0.0815).

Figure C.5: Predicted probabilities of emigration by wage residuals and different levels of differences in the standard deviations.

Germany



Source: OLS estimates from Table 7, column 2 and Polish LFS for the estimation of residual wages using a regression of log hourly wages on years of education, five age categories, marital status, gender and region of residence estimated for each year from 2004 to 2007. The difference in returns is either the average difference between the residual variation in Poland and in the destination country in the sample (Panel A: -0.0815), the median difference in the sample (Panel B: -0.074), the difference is zero (Panel C) or is at the same average difference level of the other destination country (Panel D: 0.094).