The Scandinavian
Journal of Economics

Scand. J. of Economics 117(2), 522–564, 2015
DOI: 10.1111/sjoe.12102

The Effect of Emigration from Poland on Polish Wages*

Christian Dustmann

University College London, London WC1H 0AX, UK
c.dustmann@ucl.ac.uk

Tommaso Frattini

University of Milan, 20122 Milan, Italy
tommaso.frattini@unimi.it

Anna Rosso

National Institute for Economic and Social Research, London SW1P 3HE, UK
a.rosso@niesr.ac.uk

Abstract

In this paper, we analyse the effect of emigration from Poland on Polish wages. Focusing on the 1998–2007 period for Poland, we use a unique dataset that contains information about household members who are currently living abroad, which allows us to develop region-specific emigration rates and to estimate the effect of emigration on wages using within-region variation. Our findings show that emigration led to a slight increase in wages for high- and medium-skilled workers, which are the two groups with the largest relative outmigration rates. Workers at the low end of the skill distribution might have experienced wage decreases.

Keywords: Impact; migration; sending countries
JEL classification: *J*31; *J*61

I. Introduction

Since the late 1990s, Poland has experienced a dramatic increase in emigration. Whereas in 1998, the share of emigrants in the overall population was about 0.50 percent, by only a decade later, it had increased to 2.3 percent.[1] Nevertheless, there is large regional variation in emigration rates, with a 2007 share of emigrants that ranges between 1 and 5.6 percent

*We gratefully acknowledge financial support from the Rockwool Foundation, and we are grateful to Torben Tranaes for his support throughout the project and for many helpful discussions. We also thank Iga Magda for her invaluable help with the data, and Bernd Fitzenberger for helpful comments. C. Dustmann acknowledges support from the NORFACE migration programme.

[1] See Table 1, based on the Polish Labour Force Survey.

across Poland's 16 provinces (see Table 2). This decade also saw a change in the composition of the emigration flow: emigrants became increasingly younger and were better educated than non-emigrants. These large increases in emigration, together with the variation in emigrant skill composition, are likely to have had a notable impact on the Polish labour market and, in particular, on the wages of those who stayed behind. It is this question that we address in this paper.

Specifically, we investigate the impact of emigration on wages over a period of 10 years (1998–2007) when emigration from Poland was at its highest. Because our dataset includes rare detailed information on emigrants and their education and age structure, it allows us to assign emigration rates to local labour markets and to determine the emigration-induced changes in skill ratios within local labour markets. We use the variation in emigration rates within Poland's regions to identify the effects of emigration on the wages of non-emigrants.

Although our paper is related to the body of literature on the impact of migration on wages, rather than concentrating, as most studies do, on the impacts on wages in the countries of destination,[2] it is part of only a small body of work that investigates the impact of emigration on the labour markets of sending countries. One reason such studies are scarce is the difficulty of obtaining information on emigrants, a problem that Aydemir and Borjas (2007) and Mishra (2007) overcome by exploiting the fact that over 95 percent of emigrants from Mexico go to the US. After first measuring the size and composition of Mexican emigrants from US censuses and wages in Mexico from Mexican censuses, these authors follow the identification strategy proposed by Borjas (2003) and correlate the wages of different skill groups in Mexico, defined in terms of age and education, to the proportion of emigrants from the same skill group in the US. Elsner (2013a) uses a similar approach to study Lithuanian emigration, but he must rely on a number of simplifying assumptions to reconstruct the size of Lithuanian emigration based on Irish and UK data.[3]

[2] See, for instance, early work by Altonji and Card (1991), Card (2001), Angrist and Kugler (2003), Borjas (2003), Dustmann *et al*. (2005), Card and Lewis (2007), Jaeger (2007), and more recent papers by D'Amuri *et al*. (2010), Glitz (2012), Manacorda *et al*. (2012), Ottaviano and Peri (2012), and Dustmann *et al*. (2013).

[3] Other papers on the labour market effects of emigration include: Hanson (2007), who compares changes in labour market outcomes between 1990 and 2000 in Mexican states with high and low historical levels of migration (measured in the 1950s); Docquier *et al*. (2014), who use an aggregate production function model to simulate the effect of immigration and emigration on wages and employment in OECD countries; and Elsner (2013b), who uses a calibrated structural model of labour demand to simulate the effect of Lithuanian emigration on the wages of non-emigrant workers.

We contribute to this body of literature by focusing on one large European country, Poland, which, although locked away behind the Iron Curtain for more than four decades, experienced a large amount of emigration from the late 1990s onwards. Rather than identifying emigrants based on census data and survey information from the destination countries, however (as did the aforementioned studies), we have access to detailed information (including age and education) on all emigrants measured in the source country, which allows precise computation of the regional distribution of emigrants in the country of origin. We are therefore able to estimate the effect of emigration on resident wages using an identification strategy that relies on regional variation in emigration, rather than on variation across skill groups, as in previous work. This allows us to estimate the overall effect of emigration on wages rather than its partial effects; see Ottaviano and Peri (2012) for an insightful discussion. The availability in the dataset of wage information for a subset of emigrants before they left the country also helps us to address the possible change in the composition in the non-emigrant population due to selective outmigration.

To better structure our empirical analysis and interpret our parameter estimates, we first present a model in which output is produced by combining capital with a constant elasticity of substitution (CES) labour composite. This model shows that wage effects are positive for skill groups in which outmigration falls above a weighted overall average along the skill distribution, and that – if capital is insufficiently mobile in the short run – the overall wage effects can be expected to be positive. Our empirical results, based on estimations using within-region variation, suggest that, overall, emigration had a positive effect on the wages of those who did not emigrate. Across skill groups, it is those in the middle of the educational distribution particularly that experienced the largest gains from emigration. The effect on the highly educated is likewise positive, but smaller, while the effect on the wages of those with a low level of education is negative, albeit mostly not significantly different from zero. This result is thus in line with emigration being more concentrated among individuals in the middle and upper parts of the educational distribution.

Because emigration from a particular region might be induced by negative wage shocks, we argue that such estimates are likely to constitute a lower bound on the effect of emigration on wages, and have therefore a meaningful interpretation. However, as a robustness check, we also develop an instrumental variable (IV) estimation strategy based on the detailed information we have available on the emigrants' destination countries. We combine these data with both the variations in economic conditions in the main destination countries (Ireland, Germany, the UK, and the US) and the large exchange rate fluctuations over this period, and we employ various strategies that exploit regional differences in destination preferences.

The structure of the paper is as follows. In Section II, we give a brief overview of Polish emigration, describe the data, and outline the emigrants' main characteristics. In Section III, we explain the theoretical model, while in Section IV, we describe our empirical strategy. In Section V, we present the ordinary least-squares (OLS) and IV results, and report several robustness checks. In Section VI, we discuss the results and present our conclusions.

II. Background, Data, and Descriptive Evidence

Recent Emigration from Poland

In the period after World War II, emigration from Poland was relatively low and mostly driven by political motives. Also, in the first decade following the fall of the Iron Curtain, Polish emigration remained quite modest because of relatively favourable economic conditions. From about 1998 onwards, however, after a slowing in GDP growth and a decrease in employment, emigration began increasing steadily until it peaked in 2007. Figure 1 and Table 1 provide more detailed information about the overall recent emigration trends from Poland, based on data from the Polish

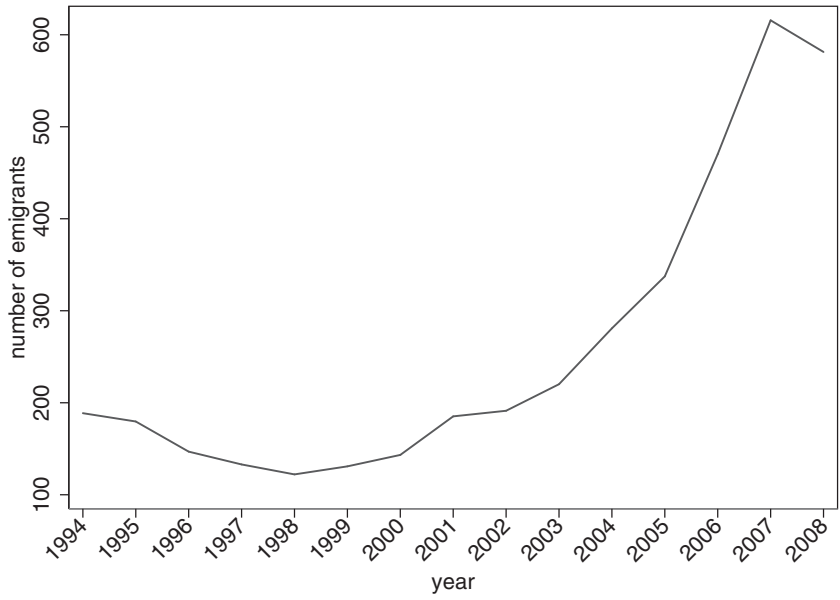

Fig. 1. Total number of emigrants, from 1994 to 2008, in thousands
Source: PLFS.
Notes: Total stock of Poles residing abroad. Average of the quarters over each year.

Table 1. *Number of Poles abroad*

	Stock (in thousands)	Change (percent)	Flow (in thousands)	Population share (percent)
1994	192.472			0.79
1995	185.389	−3.7	−7.083	0.74
1996	153.227	−17.3	−32.162	0.61
1997	139.805	−8.8	−13.422	0.55
1998	127.515	−8.8	−12.290	0.50
1999	133.247	4.5	5.733	0.51
2000	146.656	10.1	13.408	0.56
2001	191.166	30.4	44.511	0.72
2002	199.418	4.3	8.251	0.76
2003	229.833	15.3	30.416	0.87
2004	288.444	25.5	58.610	1.08
2005	343.884	19.2	55.440	1.29
2006	477.664	38.9	133.780	1.77
2007	626.927	31.2	149.263	2.29
2008	590.658	−5.8	−36.269	2.17

Source: PLFS.

Notes: In the first column, we report the stock of working-age (15–65) emigrants in each year; in the second column, the percentage change in the stock with respect to the previous year; in the third column, the flow of emigrants, given by the difference in the stock of the year with the previous year. The last column is the share of emigrants in the total working-age (15–65) population.

Labour Force Survey (PLFS), where observations are weighted using population weights (see below for details). As the figure shows, the stock of emigrants nearly quintupled between 1998 and 2007, from just above 100,000 in 1998 to over 600,000 in 2007, but decreased slightly from 2007 to 2008 because of the global economic crisis, which severely affected the main destination countries.

In 2004, Poland became a member of the European Union (EU), which gave its citizens the right to travel freely across all EU member states. In addition, the UK, Sweden, and Ireland allowed Polish citizens full access to their labour markets, while the other EU countries implemented a seven-year transition arrangement under which Poles were refused the right to work. This constraint was not strictly imposed by all countries, however; for instance, Germany (on a case-by-case basis) gave many Poles access to its labour market, which led to annual increases in the number of emigrants of between 20 and 40 percent in the years after 2004; see Kaczmarczyk and Okólski (2008) and Kaczmarczyk *et al.* (2009) for details on post-accession Polish emigration.

Data

Polish Labour Force Survey. The main dataset for our analysis is the PLFS, a rotating quarterly panel of about 15,000 households, or 50,000

individuals per quarter, conducted by the Polish Central Statistical Office (GUS) in all Poland's 16 provinces (*voivodeships*).[4] This survey covers all individuals aged 15 and above who are living in the same household, and each household is interviewed four times: in two initial consecutive quarters and then again in two consecutive quarters after a gap of two quarters. Thus, the entire interview period spans 1.5 years. We focus on the data for the 1998–2007 period.

The PLFS provides information on demographic, personal, and household characteristics of all the individuals interviewed, including age, education, current and past region of residence, country of birth, and number of children. It also collects detailed information on the economic activity of each household member during the week preceding the interview, including employment status, work arrangements, occupation, industry, and monthly net wages. In addition, and most important for our analysis, it gathers detailed demographic information – age, education level, region of origin, relationship with other household members, and country of present residence – for individuals who are part of the household but who have been residing abroad for more than three months. This information allows us to construct a comprehensive measure of outmigration. The survey also provides population weights for the resident population, which we employ throughout the analysis and also use to reconstruct population weights for emigrants. We detail this procedure in Appendix A.

Datasets from Other Countries. In addition to the PLFS, for some parts of the analysis, we draw on microdata for Germany, the UK, and the US, and aggregated data for Ireland (the four main countries of destination for Polish emigrants). In particular, we use these data to cross-check the validity of the emigration measures in the PLFS (see below) and to construct our instrumental variables (see Section IV), which are based on wage growth in the destination countries.

Information for Germany comes from the Institute for Employment Research (IAB) Employment History Data, a dataset of administrative social security records available for 1975–2007. These data encompass all individuals covered by the social security system, which is about 80 percent of the German workforce, including all workers who are subject to social security contributions (excluding the self-employed and public employees). Because the database gives no information on country of birth, however, immigrants in this dataset can only be identified based on their nationality. For the UK, we rely on the UK Labour Force Survey (LFS), a quarterly rotating panel survey available in its current format since 1992, which contains rich demographic and labour market information, including gross

[4] See Appendix A for a more detailed description of the sample used.

wages, country of birth, and years since migration. The US data come from the Integrated Public Use Microdata Series (IPUMS) of the March Current Population Survey (CPS), an integrated dataset covering 48 years (1962–2009) of the March CPS. The CPS is a monthly household survey that gathers information on labour market status and demographics, including country of birth and years since migration. Because neither the Irish Labour Force Survey nor any other Irish micro dataset contain information on wages, we use aggregate wage information for Ireland provided by the Irish Central Statistical Office (2008), which reports weekly earnings by industrial sector, gender, and type of employee. These data are based on the Earnings Hours and Employment Costs Survey (EHECS), a quarterly survey that covers all sectors of the economy other than Agriculture, Forestry, and Fishing (NACE 5–96) using a sample of 7,500 enterprises that report information on the number of employees, hours, earnings, and bonuses in that quarter.

Sample and Variables Construction

We use the PLFS data for 1998–2007 to construct the two key variables for our analysis: (i) emigration rates, by region and time period; (ii) non-emigrant wages, by region, time period, and educational group. We restrict our analysis to the age group between 15 and 65 years.

Emigration Rates. One strength of the PLFS is that it reports information on household members who are emigrants. Specifically, when a household member is not present, another member of the household is asked about the person's whereabouts. If the individual emigrated abroad more than three months earlier,[5] detailed information on age, education, country of emigration, and the individual's role in the household is collected in a separate questionnaire. This information, from which we construct our emigration rates, allows direct measurement of emigrant's individual characteristics, a major advantage over other studies that rely on destination country information to characterize emigrants. As we show in Table A1 in Appendix A, our sample comprises, on average, about 112 emigrants per region in every year, corresponding to about 10,000 individuals. In 857 cases, we observe individuals who were originally in the country but emigrated over the sampling period. For this subsample of individuals, we have a full set of information, not only standard demographics but also their wage and occupation in Poland before emigration. We use this

[5] Individuals abroad for a period of less than three months are not recorded as emigrants and cannot be separately identified in the data.

information in Section I to analyse selection patterns among emigrants, based on a comparison of their residual wages with those of non-emigrants.

However, one drawback of computing emigration rates based on these data is that such construction omits emigrants who lived in single households, as well as households in which everybody emigrates at the same time. Although, theoretically, these omissions could potentially lead to undercounting, we do not anticipate that they will pose a serious problem statistically.[6] First, the percentage of people actually living in single households in Poland is relatively small – between 8 and 9 percent, on average (in contrast to about 18 percent in the UK in 2007 and 27 percent in the US) – and this has remained fairly constant over the years, and is similar across regions. Single households are also far more frequent among the elderly: over 15 percent for the 50–64 age group versus about 8 percent for the 40–50 age group, and less than 7 percent for the 25–40 age group, which accounts for about half of all emigrants. Additionally, as reported in Section V, we also perform robustness checks in which we reconstruct the share of emigrants in the regional population by assuming that within groups defined by year, region, age, and education, the share of single households in the emigrant population corresponds to the observed share of single households in the non-emigrant population (see Appendix A for details). Finally, recent Polish emigration does not seem to be characterized by large family migration: for instance, Drinkwater *et al.* (2009) document that only about 6 percent of post-EU accession immigrants to the UK (of which Poles constitute the vast majority) brought dependants with them.

In any case, to directly assess the reliability of emigration data computed on the basis of the PLFS, we have also compared the destination-country data on trends in immigrant inflows into each country with PLFS data on trends in emigration to that particular country. As we show in Appendix A, this comparison suggests that the emigration data we are using are quite accurate.

The share of emigrants in the total working-age (15–65) Polish population showed a dramatic increase between 1998 and 2007, from 0.50 to 2.29 percent, respectively.[7] At the same time, as illustrated in Table 2, there was substantial variation in emigration rates across the different regions and into the different destination countries. For some regions, the share of emigrants over the working-age population increased more than tenfold between 1998 and 2007 (Lower Silesian), while for other regions it increased by less than 80 percent (e.g., Podlaskie). The 2007 share of

[6] In Section V, we present an interpretation of our regression results that is robust to systematic undercounting of emigrants.

[7] The emigration share is computed as the number of emigrants at time t over the working-age population (emigrants + residents) in the same year t.

Table 2. *Regional variation*

	Share of emigrants		% Germany		% US		% UK		% Ireland	
Regions	1998	2007	1998	2007	1998	2007	1998	2007	2001	2007
Lower Silesian	0.2%	2.8%	40%	26%	12%	2%	14%	37%	2%	15%
Kuyavian-Pomeranian	0.2%	1.8%	52%	12%	0%	4%	0%	43%	0%	16%
Lublin	0.7%	3.1%	11%	10%	23%	5%	1%	37%	0%	8%
Lubusz	0.4%	2.1%	55%	35%	0%	2%	0%	21%	0%	19%
Lódkie	0.2%	1.4%	22%	11%	14%	5%	17%	46%	0%	9%
Lesser Poland	1.5%	3.5%	18%	15%	41%	12%	4%	29%	1%	10%
Masovian	0.4%	0.9%	20%	0%	36%	6%	6%	54%	10%	10%
Opole	0.7%	3.6%	86%	39%	8%	0%	0%	9%	0%	12%
Subcarpathian	1.7%	5.6%	7%	10%	46%	19%	3%	22%	0%	12%
Podlaskie	1.7%	3.1%	14%	17%	49%	16%	5%	34%	0%	4%
Pomeranian	0.4%	2.1%	50%	22%	4%	0%	14%	34%	0%	18%
Silesian	0.2%	1.5%	51%	17%	5%	2%	12%	39%	0%	7%
Swietokrzyskie	0.5%	3.6%	46%	20%	9%	3%	0%	34%	0%	9%
Warmian-Masurian	0.5%	2.1%	55%	22%	6%	2%	4%	34%	3%	13%
Greater Poland	0.2%	1.6%	67%	18%	9%	1%	0%	28%	0%	24%
West Pomeranian	0.3%	2.5%	38%	16%	9%	2%	0%	29%	0%	10%
Poland	0.5%	2.4%	27%	18%	29%	6%	5%	31%	1%	12%

Source: PLFS.

Notes: The first two columns report the share of working-age (15–65) emigrants to the total working-age population in each region in 1998 and 2007. The remaining columns report the percentage of emigrants in each region for Germany, the US, the UK, and Ireland in 1998 and 2007.

emigrants ranges between 0.9 percent (Masovian) and almost 6 percent (Subcarpathian). As the table also shows, the destination countries have changed over the period. Whereas Germany was the main destination in 1997, absorbing about 27 percent of the Polish emigrant population, the largest destination country in 2007 was the UK (with 31 percent of all emigrants). Nevertheless, there is again some substantial variation across regions in the destinations chosen by emigrants.

Wages. The wage measure available in the PLFS is monthly net wages (i.e., gross wages after deduction of income taxes and social security taxes). For the construction of our wage variable (which we compute by region/year), we pool all quarters within a year, restrict the sample to the working-age population, and drop the top and bottom wage percentiles to eliminate outliers. We also eliminate all individuals who changed their migration status during the survey period, so that regional mean wages within a calendar year are always computed for the non-emigrant population only; this minimizes the changes in wages resulting from changes in sample composition. Over the period considered, real net wages increased, on average, by 1.4 percent per year.

The tax system in Poland is progressive and the tax schedule, although it varies over time, is constant across regions. Therefore, based on information about the taxation rules and the information available on each household (see Appendix A for details), we compute gross wages and then re-estimate our model to check the robustness of our results. Further, the response rate to the wage question in the PLFS decreases over our observation window, with non-response being higher among the highly educated. Hence, to check whether this decrease affects our results, we correct wages by imputing them for those who report being employed but do not report their salary (see Appendix A for the procedure used).

Descriptive Evidence

Emigrants and Non-Emigrants. How do emigrants differ from non-emigrants? In Table 3, we report the average characteristics of emigrants and non-emigrants for the years 1998 and 2007. The figures in the table show that emigrants in both years are substantially younger than non-emigrants, with the average age for emigrants decreasing by about two years between 1998 and 2007. When education level is defined as low, intermediate, or high, based on individual qualifications, emigrants are also far better educated. Low education refers to individuals with at most a lower secondary education, or up to eight years of schooling; intermediate education refers to those with a secondary education, or between nine and 13 years of schooling; high education refers to individuals with

Table 3. *Average age, gender ratio, and education in 1998 and 2007 for non-emigrants and emigrants*

	Total population		Emigrants	
	1998	2007	1998	2007
Age	38.1	38.6	34.0	32.3
Percentage of females	51%	51%	42%	34%
Education				
Percentage of low	29%	14%	12%	5%
Percentage of intermediate	60%	67%	74%	76%
Percentage of high	11%	19%	13%	20%
Percentage of those aged 25–40 years	30%	32%	47%	54%

Source: PLFS.

Notes: Entries are the average age, percentage of females, educational distribution, and the share of individuals aged 25–40 for the total population and emigrants of working age (15–65) for both sexes in 1998 and 2007. Low education refers to individuals who have at most lower secondary education, or up to eight years of schooling. Intermediate education refers to individuals with secondary education, or between nine and 13 years of schooling. High education refers to individuals with post-secondary or tertiary education, or more than 13 years of schooling.

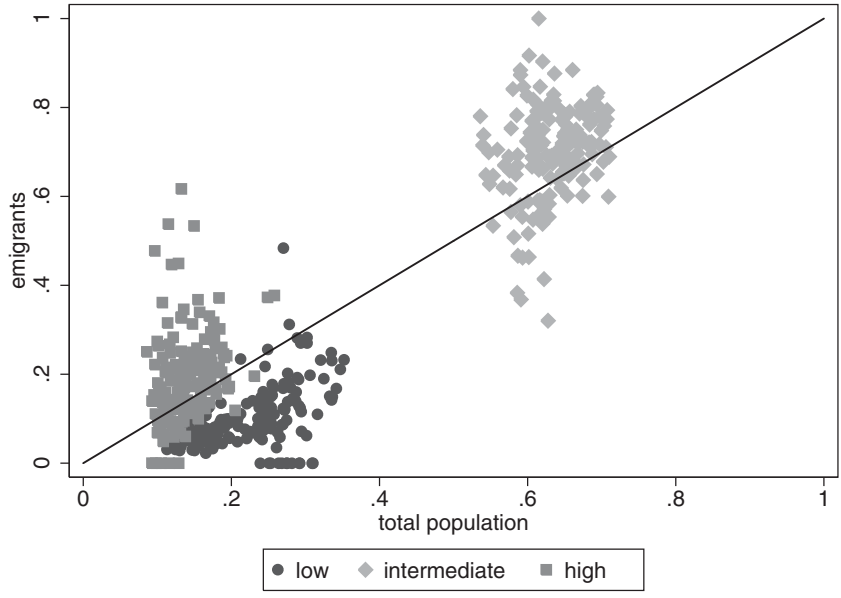

Fig. 2. Emigrants and total population: shares in each education group
Source: PLFS.
Notes: The figure plots for each region and year the proportion of working-age (15–65) emigrants in each education group versus the proportion of total working-age population in the same education group. Years 1998–2007.

post-secondary or tertiary education, or more than 13 years of schooling.[8] For both 1998 and 2007, the fraction of individuals with a low education is lower in the population of emigrants, while the fraction of those with an intermediate education is higher. The overall share of individuals with a low education decreased substantially between 1998 and 2007 for both emigrants and non-emigrants, with the drop being even larger for emigrants. These figures suggest that emigrants are over-represented among the intermediate- and high-education groups but under-represented in the low-education group.

Are these numbers similar for the different regions and across time periods? We answer this question graphically in Figure 2, which (for all years and all regions) plots the share of each education group in the emigrant population against the share of each education group in the overall population. If the skill composition of the emigrant population were identical to that of the overall population, then all dots would lie on the 45 degree line. As the figure clearly shows, however, this is not the case. For most

[8] See Appendix A for a detailed explanation of the original classification in the PLFS.

region–year pairs, the share of those in the intermediate-education group – and to a lesser extent, those in the high-education group – is higher among emigrants than among the overall population. In contrast, the share of individuals with a low education is clearly lower among emigrants than in the overall population. These numbers suggest that emigration led to a decrease in the share of the population with intermediate and high education but to a relative increase in the share of the less educated. We discuss the expected consequences of this fact on non-emigrant wages in the following subsection.

Destination Countries. How, then, are emigrants to the different destination countries selected along the education distribution? In the first column of Table 4, we report the share of Polish emigrants living in Germany, the UK, the US, and Ireland,[9] as well as the overall number living in any EU27 country,[10] for the years 1998 and 2007. The table reports both figures for all Polish emigrants, and for those who emigrated within the last year (recent emigrants). In 1998, almost one-third of all Polish emigrants lived in the US, just under 30 percent in Germany, and only 5 percent in the UK. The new flows of emigrants, however, were mostly directed toward Germany (36 percent), and to a lesser extent the US (15 percent), with only 6 percent of new emigrants going to the UK, and no emigration to Ireland. By 2007, in contrast, the situation was reversed: one-third of Polish emigrants were now living in the UK, 18 percent in Germany, 12 percent in Ireland, and only 6 percent in the US. This shift reflects a sharp change in the destination of emigration flows. In 2007, 37 percent of new Polish emigrants chose the UK as a destination, 12 percent chose Ireland, 16 percent chose Germany, and only 3 percent moved to the US. In that same year, 88 percent of the entire population of new Polish emigrants moved to EU countries, which by 2007 accounted for 84 percent of all Polish emigrants, up from 55 percent in 1998.

However, the destination countries do differ greatly in the composition of their Polish immigrant population. In Columns 2–4 of Table 4, we report the distribution of immigrants across education groups in each destination country, and in Column 5, we show the average age of emigrants in the different countries. Emigrants to Germany and the US are older and less educated, while emigrants to the UK and Ireland are far younger, with a higher share of those with intermediate or high education. Although the average age of emigrants in the US has remained stable over the years, emigrants to Germany have become older, especially compared to the average age of the total emigrant population.

[9] According to the PLFS, there were no emigrants to Ireland in 1998.

[10] In Table A2 of Appendix A, we break down the percentage of emigrants residing in each of the European countries.

Table 4. *Emigrant education by destination country*

		Education			
	% total emigrants	Low	Intermediate	High	Average age
Germany					
All emigrants					
1998	27%	11%	78%	11%	33
2007	18%	7%	82%	11%	37
Recent emigrants					
1998	36%	11%	77%	12%	32
2007	16%	7%	80%	12%	35
UK					
All emigrants					
1998	5%	10%	67%	23%	26
2007	31%	4%	71%	26%	29
Recent emigrants					
1998	6%	16%	75%	9%	25
2007	37%	4%	71%	25%	28
US					
All emigrants					
1998	29%	16%	74%	10%	39
2007	6%	3%	77%	19%	40
Recent emigrants					
1998	15%	13%	72%	15%	32
2007	3%	2%	71%	26%	34
Ireland					
All emigrants					
1998	0%	N/A	N/A	N/A	N/A
2007	12%	2%	72%	26%	30
Recent emigrants					
1998	0%	N/A	N/A	N/A	N/A
2007	12%	2%	70%	28%	29
Europe					
All emigrants					
1998	55%	11%	78%	11%	31
2007	84%	5%	77%	17%	32
Recent emigrants					
1998	73%	10%	80%	10%	30
2007	88%	5%	76%	19%	31

Source: PLFS.

Notes: In Column 1 we report the share of all working-age (15–65) emigrants and the share of recent emigrants (those who emigrated within the last year) in the total working-age population for Germany, the UK, the US, Ireland, and Europe (EU27) in 1998 and 2007. In Columns 2–4, we report the distribution of education for each group in 1998 and 2007, and in the last column we report the average age of each group in 1998 and 2007.

III. Theoretical Framework

Drawing on work by Dustmann *et al.* (2013), we next develop a model that helps us to interpret the parameters estimated below. Assume an economy with one output good (denoted by y) whose price is normalized to 1, and with multiple labour types, $i = 1, \ldots, L$. The economy is described by a

nested CES production function, which produces output y by combining a labour composite H with capital K,

$$y = [\gamma H^s + (1-\gamma)K^s]^{1/s}, \tag{1}$$

where H is a CES aggregate of the different labour types l_i, $H = (\sum_i \alpha_i l_i^{\sigma})^{1/\sigma}$, and the productivity of each labour type i is determined by α_i. Accordingly, $\sigma \leq 1$ defines the elasticity of substitution between labour types, γ defines the relative productivity of labour and capital, and $s \leq 1$ defines the elasticity of substitution between capital and labour.

Emigrant and non-emigrant labour of the same type are assumed to be perfect substitutes and equally productive, so that non-emigrant labour of type i, l_i, is the difference between labour before migration l_i^0 and emigrant labour, l_i^1: $l_i = l_i^0 - l_i^1$. Market clearing requires $l_i = n_i$ for all i, where n_i is the supply of labour of the ith type. The labour supply n_i is then the difference between labour supply in the particular skill group before emigration n_i^0 and emigrants n_i^1, so that $n_i = n_i^0 - n_i^1$. It follows that $n_i = N(\pi_i^0 - \pi_i^1 m)$, where $N = \sum_i n_i^0$ is total (pre-migration) labour supply, $\pi_i^0 = n_i^0/N$ is the fraction of total labour supply of the ith type, $\pi_i^1 = n_i^1/\sum_j n_j^1$ is the fraction of emigrant labour of the ith type, and $m = \sum_j n_j^1/N$ is the ratio of emigrants to the total (pre-migration) labour force.

In this set-up, it can be shown that the equilibrium change in non-emigrants' log wages in response to changes in the ratio of emigrants to total population is given by

$$\left.\frac{\mathrm{d}\ln w_i}{\mathrm{d}m}\right|_{m=0} = (1-\sigma)\left(\frac{\pi_i^1}{\pi_i^0} - \phi\sum \omega_j \frac{\pi_j^1}{\pi_j^0}\right), \tag{2}$$

where ω_i is the contribution of the ith type to the labour aggregate H, with $\sum_i \omega_i = 1$, and $\phi \leq 1$ is a parameter that depends on capital mobility, capital–labour substitutability, and the labour share. Importantly, $\phi = 1$ when capital is perfectly mobile, and $\phi < 1$ if capital is immobile (for more details, see Dustmann *et al.*, 2013). When $\phi = 1$, the expression in parentheses is the difference between the relative density of emigrants in the pre-migration population at skill type i, π_i^1/π_i^0, and the (weighted) relative density of emigrants across all skill types. As $\sigma \leq 1$, the wage of any skill type is increased by emigration if and only if the intensity of emigration in that skill group (the first term in parentheses) exceeds an appropriately weighted average of emigration intensity across all skill types (the second term in parentheses). If the distribution of skill types in the emigrant outflow exactly matches that in the total labour force (before emigration), then $\pi_i^0 = \pi_i^1$ for all i and the effect on wages everywhere is zero. Further, if capital is imperfectly mobile ($\phi < 1$), even emigration that

matches the pre-migration labour force in composition will result in wage gains because $\pi_i^1/\pi_i^0 > \phi \sum \omega_j(\pi_j^1/\pi_j^0)$.

The effect of emigration on the mean wages of those who do not emigrate $\sum_i w_i \pi_i^0$ is

$$\frac{\mathrm{d}\sum_i w_i \pi_i^0}{\mathrm{d}m} = (\sigma - 1)(1 - \phi)\bar{w}^0 \sum_i \omega_i \frac{\pi_i^1}{\pi_i^0} \geq 0, \tag{3}$$

where $\bar{w}^0$ is the mean wage before emigration. It follows from equation (3) that if capital is perfectly mobile ($\phi = 1$), then emigration has no effect on the mean wages of non-emigrants. However, if capital is not perfectly mobile ($\phi \leq 1$), then emigration will have a positive effect on mean wages of non-emigrants.

IV. Empirical Implementation

In our empirical implementation, we define skill groups based on education level, and estimate the following regression equation:

$$\ln w_{irt} = a_{ir} + b_{it} + c_i X_{rt} + \beta_i m_{rt} + \varepsilon_{irt}. \tag{4}$$

Here, $\ln w_{irt}$ are log mean wages of the non-emigrant population in skill group i, period t, and region r, a_{ir} and b_{it} are region and time dummies, collecting terms that vary across regions and over time, and X_{rt} controls for changes in the age, gender, and skill composition of the overall labour force. Equation (4) can be obtained by a Taylor approximation of the first-order conditions of our model around $m = 0$, and using equation (2). The parameter β_i corresponds to the term $(\sigma - 1)[\pi_i^1/\pi_i^0 - \phi \sum \omega_j(\pi_j^1/\pi_j^0)]$ given above, and measures the effect of emigration on the wages of skill groups i. Our model thus provides a clear-cut prediction for our parameter estimates: when we regress wages for a particular skill group on the fraction of emigrants to the overall workforce, m_{rt}, the sign of this parameter estimate is positive if emigrants are more densely represented in that skill group than the total population (emigrants and residents). Additionally, the larger this estimate, the smaller the short-run supply elasticity of capital. Finally, it follows from equation (3) that emigration will have a positive effect on the average wages of those who stay behind if the elasticity of capital supply is smaller than 1 so that capital is not perfectly mobile (at least in the short term). Note that our estimation strategy does not require us to break emigrants down into different skill groups.

We measure m_{rt} as the ratio of working-age emigrants in a particular region at a particular point in time to the total regional working-age population before emigration: $m_{rt} = Emigrants_{rt}/(Emigrants_{rt} + Residents_{rt})$. The vector X_{rt} collects additional control variables about the resident

Table 5. *Descriptive statistics*

	Non-emigrants		Emigrants	
Variables	Mean	Std dev.	Mean	Std dev.
Share of emigrants	–	–	1.2%	1.0%
Percentage of females	51%	1%	40%	11%
Age	38.3	0.5	33.0	2.4
Intermediate/low education	3.0	1.1	12.8	10.2
High/low education	0.7	0.3	3.0	3.3
Net wages				
Log average	6.99	0.06	6.95	0.34
Log average low education	6.74	0.09	6.40	1.72
Log average intermediate education	6.94	0.59	6.55	1.45
Log average high education	7.23	0.08	6.62	1.39

Source: PLFS.

Notes: We report pooled means and standard deviations for all regions and years (1998–2007). Entries are the percentage of females, the age, share of individuals with intermediate and high education over those with low education, real net average wage, and real net wages by education group for non-emigrants. For emigrants, we also report the share of emigrants over the total working-age population. For emigrants, wages are wages in Poland before emigration. Real wages are at 2008 prices. Non-emigrants and emigrants in the working-age population (15–65).

regional population, which include the average age, the share of women, the ratios of the number of individuals with high and intermediate education to the number of individuals with low education in the region, and the logarithm of the resident regional population. We give details of these variables in Table 5.

We estimate equation (4) by conditioning on region-specific fixed effects, thereby effectively identifying the impact of emigration on wages through variation in the emigration share (m_{rt}) within regions and over time, controlling also for year fixed effects. For regions, we use all 16 Polish *voivodeships* and for time, the years 1998–2007 (see Table 2). The resulting data include 160 observations for each skill group.

Internal Migration and Composition Effects

Internal Migration. If regions that experience high international emigration are also receiving internal immigrants, it could offset the effects of international emigration and lead to an underestimation of the effect of emigration on wages. If, instead, the same regions that experience high international emigration also experience emigration to other Polish regions, it could lead to overestimation of the effect of international emigration on regional wages (for a related discussion, see Borjas *et al.*, 1996, 1997). Because the PLFS, since 2001, reports information on region of residence one year before the interview, we can use these data to check the degree

of internal migration across the different Polish regions and the nature of its association with international migration.[11]

Overall, internal mobility in Poland across regions is low and decreasing over time; for instance, in 2001, 0.24 percent of the population reported living in a different region than in the previous year, and this share decreased to 0.12 percent in 2007. To check whether these internal movements are correlated with international emigration, we regress the share of internal migrants in the total regional population on the share of international emigrants, controlling for region fixed effects and time dummies. The resulting estimate is small, negative, and not statistically significant (we estimate a coefficient of -0.041 with a standard error of 0.027). Likewise, regressing the share of internal migrants in each region and year on the share of international migrants by skill group (controlling for year and regional dummies) produces estimates that are not significantly different from zero.[12]

We also run further regressions along the lines of Card and DiNardo (2000) to check whether emigration does indeed affect the proportion of the population in different skill groups. However, we find no evidence that the internal mobility decisions of individuals in a skill group are affected by the international emigration of individuals in the same skill group (results are available on request).

Composition Effects. A further source of concern is the possibility that emigrants are not a random sample of the regional population within each skill group i.[13] If migrants within skill group i are positively (negatively) selected, then average wages for Polish residents in skill group i could decrease (increase) after emigration purely as a result of a composition effect.[14] To check for such selection, we compare the log-wage residuals of non-emigrants versus emigrants using the sample of 857 emigrants for

[11] Note also that the regions we use are relatively large, so that internal migrations that might occur may take place within these regions.

[12] Our estimated coefficients (standard errors) are -0.023 (0.028) for the low-education group, -0.021 (0.027) for the intermediate-education group, and 0.006 (0.06) for the high-education group.

[13] For example, see Chiquiar and Hanson (2005) and Fernández-Huertas Moraga (2011) for recent studies on emigrants' self-selection, and see Hatton and Williamson (2006) for an historical overview.

[14] A related concern is that emigration might affect the labour market decisions of the other members of the households, and thus change the labour force composition. Some studies have indeed shown that in several developing countries international emigration affects the labour supply of other household members who stay behind; for example, see Amuedo-Dorantes and Pozo (2006) for Mexico, and see Antman (2013) for a recent review of this body of literature. However, the effects estimated in that literature are generally small, and we do not address this in our analysis.

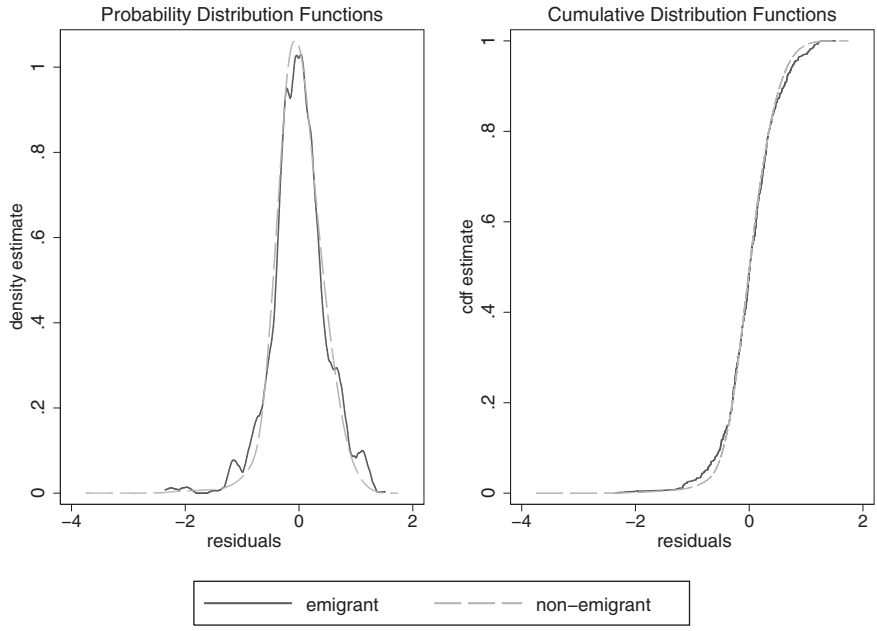

Fig. 3. Distribution of wage residuals for emigrants and non-emigrants
Source: PLFS.
Notes: The figure on the left plots the pdf estimates of residuals for emigrants and non-emigrants. The estimates are obtained using a Gaussian kernel with optimal bandwidth. The figure on the right plots the cumulative distribution functions of residuals for emigrants and non-emigrants.

which we have pre-emigration wage data (see Section II).[15] The overall mean difference in residual wages between emigrants and non-emigrants is 0.0171, and it is not significantly different from zero (standard errors = 0.014). Neither are the mean differences by education group: the difference in wage residuals for those with low education is 0.032 (standard error = 0.0584), whereas it is 0.021 (standard error = 0.0157) for those with an intermediate level of education and −0.005 (standard error = 0.035) for the highly educated.

In Figure 3, we plot the kernel density estimates of the residual distribution for emigrants and non-emigrants (left panel) and the cumulative distribution functions of the two groups (right panel). Visual inspection suggests that these distributions are very similar for the two groups. The *D*-statistics on a Kolmogorov–Smirnov test on the equality of the empirical

[15] We compute residuals from a weighted log-wage regression on education (three categories): age; age squared; region and year dummies.

distribution functions is 0.039 (p-value = 0.147), so we cannot reject the null hypothesis that the two samples are drawn from the same distribution.[16]

Non-Random Emigration and OLS as Lower Bound Estimates

A further potential problem with specification (4) is that emigration choices might not be random. That is, although region fixed effects account for permanent regional differences – and therefore also for the fact that, for instance, emigration might be higher from rural or traditionally less wealthy regions – even after they are controlled for, region-specific shocks affecting the wages of skill group i in year t (ε_{irt}) could be correlated with regional emigration flows in the same year. In that case, OLS estimates would be biased.

If, as seems plausible, emigration is higher from regions that experience negative wage shocks, then this association might induce a spurious negative correlation between emigration and wage growth that would lead to a negative bias in the OLS estimate of the effect of emigration on mean wages. Hence, the OLS estimator provides a lower bound for the actual effect of emigration on mean wages. Additionally, the OLS estimates are also a lower bound for the effect of emigration on the wages of each skill group i, as long as skill-specific shocks are positively correlated within regions in every year. We test this assumption by running separate pairwise regressions of regional wage growth rates for each skill group on the wage growth of all other skill groups, controlling for year dummies. In all cases, we find that the growth rates of wages for all skill groups are positively correlated within regions, although the estimated coefficients are not significant for the correlation between the wages of the individuals in the low- and high-education groups.[17]

Instrumental Variables Estimation

As we explain above, under plausible assumptions we can interpret our OLS estimates on the effect of emigration on the wages of non-emigrants as lower bounds. However, the bias of the OLS estimates could go in the opposite direction if other unobservable factors are simultaneously affecting

[16] The two-sample Kolmogorov–Smirnov test is a non-parametric test for whether the two samples are drawn from the same distribution. The D-statistics measures the distance between the cumulative distribution functions of the two samples and the null hypothesis is that the samples are drawn from the same distribution.

[17] We estimate a coefficient (standard error) of 0.898 (0.257) for the regression of low-skilled wages on intermediate-skilled wages; 0.196 (0.145) for the regression of low-skilled wages on high-skilled wages; 0.303 (0.148) for the regression of high-skilled wages on intermediate-skilled wages.

positively emigration flows and wage growth. Although we do not believe that such a scenario is plausible, we nevertheless provide robustness checks by using an IV strategy. This requires an instrument, or set of instruments, that is correlated with m_{rt}, the ratio of emigrants over the total population in region r at time t, but uncorrelated with ε_{irt}, the economic shock hitting region r at time t, conditional on time–region dummies and the set of individual characteristics included.

Such instruments are difficult to find in the study of the effects of emigration on wages. Because Polish emigration before 1997 was very low, we cannot construct an instrument equivalent to that in the literature that estimates the effect of immigration on native outcomes based on spatial variation, which is based on past settlement patterns of immigrants (e.g., Altonji and Card, 1991; Card, 2001; Cortes, 2008; Frattini, 2010; Lewis, 2008; Bianchi *et al.*, 2012; Dustmann *et al.*, 2013). Instead, we adopt here an IV strategy based on economic shocks to destination countries, shocks that are likely to influence emigration (testable) while being uncorrelated with the shocks to a particular Polish region (our identifying assumption).

Our main measure of shocks to destination countries is the annual growth rate of real wages below the 40th percentile of the wage distribution in a set of potential destination countries.[18] We allow the effect of these shocks on the probability of migration to differ across regions through regional heterogeneity in migration costs to each potential destination country. One reason why migration costs to each country c might differ across regions are differences in historical ties with the destination country c (e.g., regions close to the German borders have stronger links with Germany; see Kraetke, 1996, 1999) or regional variation in the relative size of the existing stock of emigrants in each destination country; for example, see Munshi (2003), McKenzie and Rapoport (2007), and Pedersen *et al.* (2008) for the importance of network effects on migration decisions. Furthermore, differences in geographical proximity, in the share of each region's population who speaks the host country's language, or in the distance from airports with international flights are all factors that may lead to cross-region heterogeneity in the pull effect of shocks from destination countries.

Empirically, we model regional variation in the destination countries' shocks in several alternative ways. First, we simply interact the shock variable with regional dummies, which requires no assumptions about the type of regional heterogeneity because this is estimated from the data. However, this flexibility comes at the cost of increasing substantially the number of instruments. Therefore, we also experiment with different ways of reducing the number of instruments. In particular, we adopt different

[18] That is, the part of the wage distribution where most Polish immigrants fall in destination countries, especially in the first years after migration (see Appendix B for more details).

criteria to select only a subset of relevant instruments, or weight each country's shock with pre-assigned continuous regional weights. In the latter case, we use two alternative weights: the inverse of the distance between each region's capital and the capital of the destination country c, or the mean share of emigrants from a region to each country over the 1998–2007 period (for every year t we construct this share excluding years $t-1$, t, and $t+1$). Finally, because the 2004 EU accession has radically changed the relative migration costs from Poland to EU and non-EU countries, and to the different countries within the EU, we distinguish between the pre- and post-accession years.

To implement our strategy, we consider the four countries to which the majority of Polish emigrants migrated over the period under consideration: Germany, Ireland, the UK, and the US – countries where about 65 percent of all Polish emigrants settled between 1998 and 2007. We provide more detail about the construction of instruments and estimation procedures in Appendix B.

V. Results

In Table 6, we report OLS estimates of β in expression (4) for average wages (Row 1) and for the wages of the different education groups (Rows 2–4). Panel A reports our baseline results, while Panels B–D report different robustness checks. Column 1 reports the results from a specification that controls only for regional fixed effects and year dummies, while Column 2 reports results when controls are added for the size of the regional population, the regional age structure, and the educational and gender composition (see Section IV for more details). All regressions refer to the years between 1998 and 2007. In Panel A, we use net wages, as reported in the survey (see Section II). The estimates in Row 1 show that emigration is associated with a higher growth of regional average wages: the estimated coefficient ranges between 0.97 in Column 1 and 1 in Column 2, and is statistically significant in both cases at the 10 percent level. Because the variation used for estimation is the change in the stock of emigrants between consecutive years, these are short-run estimates. As pointed out in our theoretical discussion in Section III, a positive overall effect of emigration is compatible with the elasticity of capital supply not being infinite, at least in the short run. In terms of magnitude, the estimates in Column 2 imply that an increase of one percentage point in the ratio of emigrants to the total population led to a 1 percent increase in average real wages.[19]

[19] See also the subsection *Interpretation and Comparison with Previous Studies*, in Section V, for an interpretation of the results in terms of elasticity and a comparison with results from other studies.

Table 6. *Effects of emigration on log mean wages, OLS*

Dependent	A Net wages		B Gross wages		C Imputed wages		D Adjusting for single households	
variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average	0.969*	0.999*	1.058*	1.100*	0.65	0.860*	0.903*	0.897*
	(0.551)	(0.558)	(0.590)	(0.596)	(0.519)	(0.518)	(0.481)	(0.532)
Low education	−1.154	−2.138	−1.664	−2.762	−1.243	−2.023	−1.684	−2.570*
	(1.510)	(1.463)	(1.783)	(1.728)	(1.336)	(1.287)	(1.329)	(1.386)
Intermediate education	1.285**	1.403**	1.463**	1.619***	1.033*	1.201**	1.047**	1.148**
	(0.569)	(0.569)	(0.619)	(0.614)	(0.559)	(0.562)	(0.493)	(0.547)
High education	1.515*	1.142	1.647*	1.254	1.247*	1.037	1.903**	1.527*
	(0.861)	(0.871)	(0.906)	(0.918)	(0.671)	(0.684)	(0.751)	(0.824)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	160	160	160	160	160	160	160	160

Notes: Entries are estimated regression coefficients of the regional ratio of emigrants over the total population on regional average log net wages and on average log wages by education groups for years 1998–2007. In each panel, we use a different measure of average, low-, intermediate-, and high-education wages. In Panel A, we use net monthly wages. In Panel B, we use gross wages. In Panel C, we impute wages for employed individuals with missing wage information. In Panel D, we adjust the share of emigrants by the share of single households in the population. All regressions include region fixed effects. "Other controls" are log regional population, mean regional age and gender, the ratio of the numbers of individuals with high and intermediate education to the number of individuals with low education. Newey–West standard errors using one lag are reported in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Over the period considered, emigration from Poland increased on average by 0.19 percentage points per year and real wages increased by about 1.7 percent per year. These estimates therefore suggest that emigration might have contributed almost 11 percent to overall wage growth.

In Rows 2–4 of Table 6, we report the results for the three different education groups. The figures in Table 3 suggest that emigration was mainly concentrated in the middle part of the educational distribution and far less at the bottom. In particular, the relative intensity of emigration (the ratio π_i^1/π_i^0 in the notation of our model in Section III) was between 0.42 and 0.31 for the low educated, between 1.22 and 1.14 for those with intermediate-level education, and between 1.2 and 1.05 for the highly educated. According to the model developed in Section III, the effect of emigration should thus be felt most by those with intermediate education because this group experiences the largest (negative) relative supply shock. The results in Rows 2–4 of Table 6 are in line with these predictions, suggesting that emigration led to an increase in wages for workers with an intermediate or high level of education but possibly depressed wages

for those with low education. However, the estimates for the low-education group are not significantly different from zero.

Overall, these results conform remarkably with the predictions of the simple model outlined in Section III, with larger gains for workers in the skill categories exposed to a larger negative supply shock. They also indicate that emigration helped overall wage growth in Poland over the period under consideration, although it might have reduced returns to capital.

The next three panels provide additional results and robustness checks that address the data limitations outlined in Section II. In Panel B, we use gross wages as the dependent variable, constructed from the information on net wages and individual characteristics (see Appendix A for details). The results are very similar to those reported in Panel A, with the estimates for the low-education group being slightly larger but not significantly different from zero. In Panel C, we account for non-response to the PLFS wage question by imputing (net) wages for those individuals for whom wage information is missing (see Appendix A for details). Again, the estimates are in line with our baseline results, although the estimated coefficients are slightly smaller. In Panel D, we report the results after correction of the emigration share measure for possible under-counting because of the single emigrant households not captured in our data. We assume that single households are as frequent among emigrants as among non-emigrants within the same age–education group in every region and year, and we re-scale the number of emigrants accordingly (see Appendix A for details). This re-scaling slightly reduces the size of all estimated coefficients relative to the baseline, with the exception of those in the low-education group, which are now slightly larger.

In Panel A of Table 7, we report our IV estimates when we use as instruments wage growth (in Polish purchasing power) in each destination country, interacted with regional dummies and dummies for the period before and after EU accession (see Appendix B for details). As before, Column 1 shows results from a specification where the only additional variables are region and year dummies, while in Column 2 we display results from a richer specification where we include all control variables. We show the first-stage statistics for our instruments in Rows 5 (partial R^2) and 6 (F-statistics for joint significance of excluded instruments) of Table 7. The partial R^2 is high, suggesting that our instrument set explains about 90 percent of the variation in emigration rates. Likewise, the F-statistics for the significance of excluded instruments is 12.3 in the basic model (Column 1) and 10.2 in the model with all controls (Column 2).[20] The IV estimates are very close to the OLS estimates, indicating that

[20] According to the tabulations of Stock and Yogo (2005), the critical value for the F-statistics when using 120 instruments is 11.31.

Table 7. *Effects of emigration on log mean wages, IV*

	A Wage shocks		B GDP growth deviation		C 1% significance		D Backward		E Forward		F Distance		G Mean share	
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Average	0.968*	1.012*	0.968*	1.014*	1.201*	1.173*	1.003*	1.015*	1.405**	1.613**	2.502	1.635	2.068*	1.996*
	(0.556)	(0.561)	(0.561)	(0.565)	(0.642)	(0.650)	(0.558)	(0.564)	(0.646)	(0.651)	(1.559)	(1.338)	(1.136)	(1.158)
Low education	−1.187	−2.154	−1.165	−2.151	−0.792	−2.287	−1.04	−2.154	−1.004	−2.404	1.454	−2.254	−0.007	−1.348
	(1.522)	(1.472)	(1.536)	(1.482)	(1.757)	(1.700)	(1.529)	(1.481)	(1.769)	(1.700)	(4.227)	(3.511)	(3.083)	(3.008)
Intermediate education	1.305**	1.433**	1.301**	1.449**	1.515**	1.652**	1.378**	1.503***	1.954***	2.218***	2.824*	2.277*	3.388***	3.561***
	(0.574)	(0.572)	(0.579)	(0.577)	(0.663)	(0.664)	(0.576)	(0.576)	(0.668)	(0.665)	(1.601)	(1.365)	(1.208)	(1.220)
High education	1.534*	1.154	1.407	1.026	2.129**	1.668	1.464*	1.021	1.558	1.036	0.869	0.736	0.887	0.356
	(0.868)	(0.876)	(0.876)	(0.883)	(1.003)	(1.015)	(0.872)	(0.882)	(1.009)	(1.014)	(2.385)	(2.091)	(1.758)	(1.796)
(Adjusted) partial R^2	0.92	0.93	0.85	0.85	0.57	0.54	0.94	0.93	0.74	0.75	0.13	0.17	0.24	0.24
F-statistics for excluded instruments	12.3	10.21	7.29	6.43	6.45	6.07	38.5	35.72	14.31	14.47	2.05	2.81	4.43	4.24
No. of excluded instruments	120	120	120	120	44	44	77	77	23	23	8	8	8	8
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	160	160	160	160	160	160	160	160	160	160	160	160	160	160

Notes: Entries are estimated regression coefficients using different instrumental variables for the regional share of emigrants over the total population in regressions of average regional log net wages on the regional emigrant shares and region and year dummies for years 1998–2007. In Panel A, the IVs are the annual growth rate of real wages below the 40th percentile in country j (US, UK, and Germany) and wages in the construction and manufacturing sector for Ireland, expressed in zloty interacted with regional dummies. In Panel B, we use the GDP growth deviation and we interact regional dummies with the deviation of the GDP growth of destination country j from the OECD mean GDP growth (in US constant dollars). In Panel C, we select instruments that are significant at 1 percent and run the first-stage regression using the selected instruments only. In Panels D and E, we select the variables using a backward (Panel D) and forward (Panel E) selection. In Panel F, we interact the mean wage growth at the 40th percentile with the inverse of the distance of the regional capital from the capital of the country of destination. In Panel G, we interact the mean regional share of emigrants in destination country j with the wage growth in the same destination country j. "Other controls" are log regional population, mean regional age and gender, the ratio of the numbers of individuals with high and intermediate education to the number of individuals with low education. Newey–West standard errors using one lag are reported in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

emigration does have a positive effect on average wages in Poland (Row 1), while having positive effects on wages of workers with a high (Row 4) and, especially, an intermediate (Row 3) level of education. The estimated effect for the group of low-educated workers (Row 2) is again negative, although imprecisely estimated. In Panel B, the instrument used is the deviation in each destination country's GDP per capita growth rate relative to the OECD mean, interacted with regional dummies and EU accession dummies. The first-stage statistics, reported in Rows 5 and 6, indicate that these instruments are weaker than the instruments used in Panel A. Nevertheless, the results from both instrument sets are very similar.

In Panels C–G, we present IV estimates where we reduce the number of instruments. In Panel C, we use as instruments the wage growth in destination countries, but consider only those region–country pairs for which wage growth is statistically significant at the 1 percent level of significance. This strategy reduces the number of instruments to 44, and leads to lower first-stage statistics, but the estimated coefficients are very close to those in Panels A and B. In Panels D and E, we show estimates from a model where the IVs are selected from the potential instruments set using backward (Panel D) or forward (Panel E) selection.[21] In both cases, the first-stage statistics increase considerably, although the number of instruments is reduced to 77 and 23, respectively. Estimates are again similar to our baseline results, although the magnitude increases slightly, especially in the case of forward selection.

In Panels F and G, we do not estimate the weights for the destination countries' wage growth, but we compute them. For the results in Panel F, we use the inverse of the distance between each region and the destination country to compute the weights; in Panel G, we use the mean (over time) of the regional emigration share to each destination country. The partial R^2 (F-statistics) is lower in both cases, ranging from 0.13 (2.05) in Panel F to 0.24 (4.43) in Panel G. Nevertheless, the estimates in Panels F and G are again similar to results in previous columns.

Overall, the IV results are remarkably stable despite the different instruments and weights used. While the weighting factors for the first set of instruments in Panels A–E are estimated, the results of Panels F and G rely on an IV strategy with pre-determined weighting factors. Despite a lower first stage, estimates are in line with those obtained using other

[21] The backward selection is an iterative procedure where we start by running a first-stage regression including all regressors, and drop the least significant regressor at each subsequent re-estimation of the model. We iterate this procedure until all included regressors are at least statistically significant at the 10 percent level. In the forward selection, we first fit a model with a constant only. We then add each potential regressor and select the one with the highest statistical significance. We iterate this procedure to add at each iteration one additional variable, as long as it is at least statistically significant at the 10 percent level.

estimation strategies, and confirm the pattern shown in the OLS results of a slight overall increase in wages, and that workers in the intermediate skills category, who experienced the largest negative supply shock, have gained most in terms of wages.

Interpretation and Comparison with Previous Studies

How do our estimates compare to those of previous studies? As noted by Bratsberg *et al.* (2014) and Bratsberg and Raaum (2012), it might be difficult to compare our estimated coefficient in equation (4) with those obtained from studies in other countries because different studies report different coefficients, and because of cross-country differences in the size of the emigrant (or immigrant) population. Here, we follow Bratsberg *et al.* (2014) and compute the elasticity of non-emigrant wages with respect to the size of the emigrant labour force. An advantage of this measure is that it is unaffected by systematic undercounting, which could lead to an overestimate of the effects of emigration, as long as undercounting is proportional to the true number (see Section II). In our case, the elasticity of non-emigrant wages with respect to emigration is given by $\partial \ln w_{irt}/\partial \ln E_{rt} = \beta m_{rt}(1 - m_{rt})$, where β corresponds to the estimated coefficient in Tables 6 and 7, m_{rt} is the share of emigrants, and E_{rt} is the size of the emigrant population.[22] Evaluated at the sample mean emigrant share $\bar{m} = 1.2$ percent, this elasticity lies between 0.01 (OLS results and baseline IV results) and 0.02 (our largest IV estimate), suggesting that a 1 percent increase in the number of emigrants increases wages by between 0.01 and 0.02 percent. These estimates compare, for instance, to an elasticity of wages to immigration in the UK of 0.03, estimated by Dustmann *et al.* (2013). As regards the elasticity of wages by skill group, our estimates imply an elasticity between -0.03 and -0.02 for wages of low-skilled workers, and elasticities for wages of intermediate- and high-skilled workers that lie between 0.02 and 0.04, and around 0.01, respectively.

We have no comparable estimates for the elasticity of wages to emigration. The estimates from the existing papers are obtained using the skill cell correlation approach, which produces estimates of the direct partial elasticity of wages to emigration (i.e., the percentage change in wages of a skill group caused by a 1 percent increase in the number of emigrants in the same skill group, holding non-emigrant labour supply, aggregate supplies, and capital constant). Nevertheless, evaluated at the sample mean

22

$$\beta = \frac{\partial \ln w_{irt}}{\partial m_{rt}} = \frac{\partial w_{irt}/w_{irt}}{\partial E_{rt}/E_{rt}} \frac{\partial E_{rt}/E_{rt}}{\partial m_{rt}} = \frac{\partial \ln w_{irt}}{\partial \ln E_{rt}} \frac{1}{m_{rt}(1 - m_{rt})},$$

as $E_{rt} = R_{rt}[m_{rt}/(1 - m_{rt})]$, and where R_{rt} is the resident non-migrant population.

emigrant/resident ratio of 10 percent and using the preferred coefficient estimate of 0.33, the results of Mishra (2007) imply a partial elasticity of 0.033,[23] while estimates of Aydemir and Borjas (2007) for Mexico over the period 1960–2000 imply an elasticity of 0.06 (obtained by evaluating their estimates of 0.8 at the mean ratio of emigrants to total population of 8.5 percent), and the results of Elsner (2013a) imply a partial elasticity of 0.03 (evaluated at the estimated coefficient of 0.665 and at the mean share of 5 percent).

VI. Discussion and Conclusions

We use the PLFS to assess the effect that emigration over the 1998–2007 period – a time of large outmigration – had on the wages of Polish workers who did not emigrate. The PLFS is unique in two aspects. First, it asks households about household members who have migrated, which allows direct measurement of the migrant population, and second, it provides information about the emigrants' key characteristics, including age and educational level. We use these data to construct region-specific emigration rates and to assess the skill composition of the emigrant population.

Our basic results suggest that the large emigration experienced by Poland over the 1998–2007 period (when the emigrant share increased from 0.5 to 2.3 percent, and in some regions up to 5.6 percent) contributed to overall wage growth, particularly for workers in the intermediate-skill group, which experienced the largest negative labour supply shock. Because of a possible correlation between region-specific wage shocks and outmigration, OLS estimates that condition on region fixed effects are potentially downward biased, allowing us to interpret our results as lower bounds. As a robustness check, we also implement an IV strategy based on labour market shocks to the various destination countries. Our IV results using different IV strategies are – overall – slightly larger than the fixed effect results and re-confirm a slight overall positive effect of emigration, with individuals in the intermediate-education group gaining most. These results are remarkably stable, and robust to various definitions of the wage variables and checks on the potential mismeasurement of regional emigration rates.

Taken together, our findings suggest that emigration from Poland over the 1998–2007 period had a slightly positive (although not always precisely estimated) effect on the average wages of those who did not emigrate. Our results are thus in line with those found by Aydemir and Borjas (2007)

[23] In the studies of Mishra (2007) and Elsner (2013a), the regressor is the ratio of emigrants (E) to non-emigrant residents (R), $p_{irt} = E_{irt}/R_{irt}$. Thus, the elasticity is equal to the estimated coefficient times the average share $\bar{p}$.

and Mishra (2007), with a different empirical strategy, for Mexican emigration. Within our theoretical framework (see Section III), this finding implies that the supply of capital was, at least in the short run, not perfectly elastic. Moreover, the impact of migration on wages for the different skill groups seems to mirror the relative negative supply shocks experienced by these skill groups through emigration; that is, the emigrants were drawn primarily from the medium and upper parts of the educational distribution in which positive wage effects are more pronounced. Not everyone gained, however. According to our point estimates, workers with a low education – the group that emigrated least and thus became relatively more abundant – experienced no wage gains and might even have experienced wage decreases, although estimates are insignificant for this skill group in almost all cases.

Appendix A: Data

Sample Extraction

Poland is divided into 16 regions: Greater Poland (*województwo wielkopolskie*), Kuyavian-Pomeranian (*województwo kujawsko-pomorskie*), Lesser Poland (*województwo małopolskie*), Łódź Voivodeship (*województwo łódzkie*), Lower Silesian (*województwo dolnośląskie*), Lublin (*województwo lubelskie*), Lubusz (*województwo lubuskie*), Masovian (*województwo mazowieckie*), Opole (*województwo opolskie*), Podlaskie (*województwo podlaskie*), Pomeranian (*województwo pomorskie*), Silesian (*województwo śląskie*), Subcarpathian (*województwo podkarpackie*), Świętokrzyskie (*województwo świętokrzyskie*), Warmian-Masurian (*województwo warmińsko-mazurskie*), and West Pomeranian (*województwo zachodniopomorskie*). In our analysis, each region is considered to be a separate labour market. The average regional labour force (active and inactive) is about 1.7 million, with the largest region in both 1998 and 2007 being Masovian, the region of Warsaw, and the smallest regions, Lubusz in 1998 and Opole in 2007. Figures A1 and A2 each show a map of Poland with the 16 provinces marked, together with the yearly average wage increase and the yearly average change in the share of emigrants between 1998 and 2007, respectively.

For our analysis, we use data from the period 1998–2007 and restrict the sample to those between 15 and 65 years old. We eliminate wage observations below the first percentile and above the 99th percentile to eliminate outliers. To avoid selection problems resulting from changes in sample composition, we compute regional average wages by year and keep the sample used to compute these wages constant by dropping all individuals

Fig. A1. Annual average increase in real wages between 1998 and 2007
Source: PLFS.

Fig. A2. Annual average increase in the share of emigrants (in percentage points) between 1998 and 2007
Source: PLFS.

who are return migrants or emigrate within the next year. Mean wages by region are calculated using the population weights provided in the survey.

The variable *education* is defined by re-coding the original variable in the survey (which was classified into nine categories) into three categories: *low education group* = all individuals who have partially or fully completed primary school, or equivalently have eight or fewer years of education (*bez wykształcenia szkolnego*, *niepełne podstawowe*, and *podstawowe*); *intermediate education group* = all those who have completed a vocational or general secondary education and have between nine and 13 years

of education (*średnie zawodowe*, *średnie ogólnokształcące*, *gimnazjum*, and *zasadnicze zawodowe*); *high education group* = all those who have a post-secondary or higher education or more than 13 years of schooling (*wyższe* and *policealne*).

Weights: Estimation Strategy

Because the dataset for emigrants reports no sampling weights, we estimate the weights for emigrants based on those we have for members of the labour force. From the information provided by the Polish Statistical Office, we know that the sampling units are households and the first stratum of the sampling procedure consists of regions. Weights are then defined on the basis of the response rate and certain other demographic variables (place of residence, gender, and age). Based on this information, we estimate the following regression for each year t, quarter q, and gender s:

$$weights_{itqs} = \sum_{r=1}^{R} \sum_{d=1}^{D} \sum_{y=1900}^{Y} \beta_{rdytqs} R_r * D_{itqs} * Y_{itqs} + U_{itqs}, \quad \text{for } \forall\, t, q, s.$$

Here, R_r are regional dummies, D_{itqs} are eight dummies for the size of the town in which the household resides, Y_{itqs} are year-of-birth dummies, and u_{itqs} is an error term. After using the above regression to estimate the weights for emigrants,[24] we use the weights obtained to compute all other emigrant information, such as total number of emigrants and share of emigrants in the total population.

Net and Gross Wages

The PLFS contains information about net monthly wages only. We reconstruct gross wages, applying Poland's three tax rates[25] to the three income brackets identified by the two tax base thresholds. Therefore, we apply the following formula:

[24] For emigrants, there is no information on the size of their town of residence before moving abroad. However, we do have information on the household the emigrant belonged to before emigration. Therefore, we assign the emigrant's town size based on this latter information.

[25] Poland has an individual taxation system, but taxpayers can decide to pool their income with the income of other people in the family. Because we do not observe the actual behaviour of households, we compute gross wages under the assumption that workers do not choose to pool earnings.

$$grosswages_{it} = \begin{cases} \frac{netwages_{it} - TC_t}{1-\tau_t^{\mathrm{L}}} & \text{if } netwages_{it} \leq x_t^{\mathrm{L,N}} \\ \frac{netwages_{it} - TC_t + x_t^{\mathrm{L,G}}(\tau_t^{\mathrm{L}} - \tau_t^{\mathrm{M}})}{1-\tau_t^{\mathrm{M}}} & \text{if } x_t^{\mathrm{L,N}} < netwages_{it} \leq x_t^{\mathrm{H,N}} \\ \frac{netwages_{it} - TC_t + x_t^{\mathrm{L,G}}(\tau_t^{\mathrm{L}} - \tau_t^{\mathrm{M}}) + x_t^{\mathrm{H,G}}(\tau_t^{\mathrm{M}} - \tau_t^{\mathrm{H}})}{1-\tau_t^{\mathrm{H}}} & \text{if } netwages_{it} > x_t^{\mathrm{H,N}} \end{cases}.$$

Here, $grosswages_{it}$ are the yearly gross wages for individual i at time t, $netwages_{it}$ are yearly net wages, TC_t are the tax credits for which individuals are eligible, τ_t^b are the tax rates, and $x_t^{b,j}$ are the tax base thresholds, where $b = \mathrm{L}$ (low), M (intermediate), H (high) and $j = \mathrm{N}$ (net), G (gross). We apply the net threshold to our data: $x_t^{\mathrm{L,N}} = x_t^{\mathrm{L,G}}(1 - \tau_t^{\mathrm{L}})$; $x_t^{\mathrm{H,N}} = x_t^{\mathrm{H,G}}(1 - \tau_t^{\mathrm{M}})$. The fiscal year in Poland corresponds to the calendar year. For each individual, we compute the yearly net wage (from the monthly net wages reported in the survey) and assign individuals to the respective tax base bracket. In 1999, a tax reform was introduced that required employees to pay their own social contributions (previously paid by employers). Hence, to make gross wages comparable across the years, we compute gross wages net of employees' social contributions. Finally, we divide the yearly gross wages by 12 in order to obtain monthly gross wages to use in the econometric analysis.

Missing Wages

Not only did the non-response rate to the PLFS wage question increase in later years, from 17 percent in 1998 to 32 percent in 2007, but the better educated are over-represented among non-responders, which might lead to underestimation or overestimation of the effect of emigration, depending on the type of selection. Hence, to check the robustness of our results, we correct for differential non-response rates across different population groups and impute wages for employed individuals with missing wage information. Under the assumption that the probability of response to the wage question depends only on observable characteristics, this procedure allows us to recover measures of regional average log wages. Specifically, the imputation procedure works as follows. First, for each year, quarter, and gender, we run separate regressions of log wages, controlling for age and education and their interaction, occupation, marital status, part-time work, whether the individual is a public sector employee, city size, and region of residence.[26] We use the coefficients estimated in these regressions to predict wages for all employees in the sample for whom wage information

[26] We control for age using dummies for 10-year age brackets; for educational level, using dummies for low, intermediate, and high education; for occupation, using dummies for each

is missing. We add an error term to the prediction, drawn from a normal distribution, with zero mean and heteroscedastic variance according to age, education, and gender. We use these wages to compute regional means to be used in the econometric analysis.

Emigrant Share

The fact that we do not observe emigrants who live in single households before emigration implies that we might be underestimating the number of Polish emigrants. However, as explained in Section II, the percentage of single households in the age range at which most migrations take place is so low that their omission is unlikely to be a serious problem, especially as we demonstrate that our data on emigration to different countries closely resemble those constructed from data sources in the receiving countries. Nevertheless, as a robustness check, we also provide estimates in which we reconstruct the share of single household emigrants using information on the share of single households among non-emigrants.

To do so, we first compute the share of individuals living in a single household in the resident population in year t, region r, age group a (using five 10-year age brackets) and education level e (using three education levels), α_{raet}. Under the assumption that the share of single households, conditional on observable characteristics, is the same among residents and emigrants, we then re-scale the number of observed emigrants in each year–region–age–education cell E^*_{raet} by $1/(1-\alpha_{raet})$ to obtain an estimate of the actual number of emigrants in that cell, $E_{raet} = E^*_{raet}/(1-\alpha_{raet})$. We then sum up the adjusted numbers of emigrants by region and year, and compute the shares to be used in the regression. Adjusting for under-counting of single households, the share of emigrants in the total working-age population is 0.7 percent in 1998 and 2.8 percent in 2007, which compares with 0.5 and 2.3 percent, respectively, without the adjustment.

Comparison of PLFS Data with Destination-Country Datasets

In order to directly assess the reliability of emigration data computed on the basis of the PLFS, we have compared the destination-country data on trends in immigrant inflows into each country with PLFS data on trends in emigration to that particular country. We focus on the three main destination countries for Polish emigrants – Germany, the UK, and the US – which alone account for over 55 percent of all Polish immigration over the years considered.

---

one-digit ISCO08 occupation group. The controls for city size are dummy variables for seven size categories, and those for region of residence are dummies for each *voivodship*.

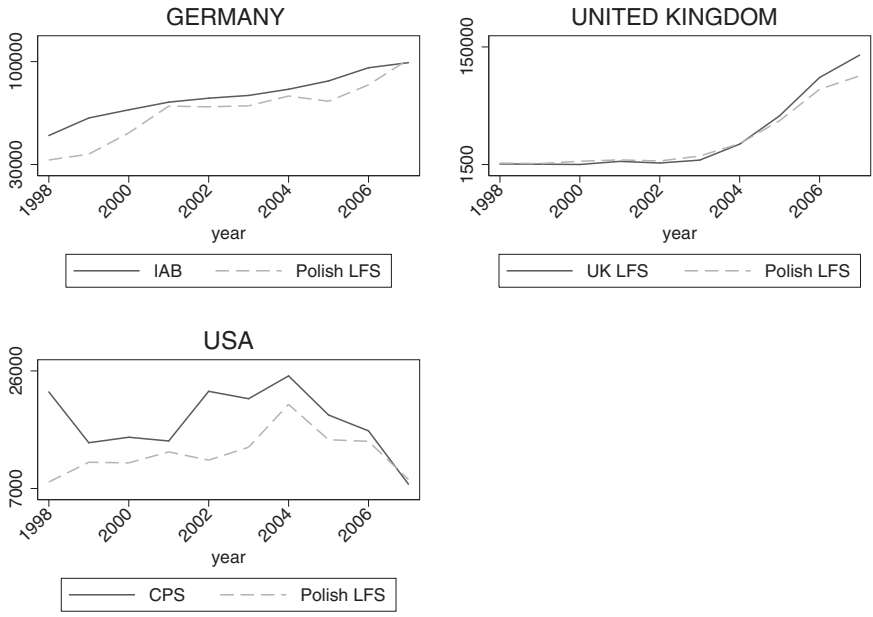

Fig. A3. Number of Poles abroad from the PLFS and other datasets: a comparison
Source: PLFS, Institute for Employment Research (IAB), UK Labour Force Survey (LFS), and Current Population Survey (CPS).
Notes: For each destination country, we report the stock of Polish immigrants between 1998 and 2007. For the UK and US, we have information on the year of arrival in the country, so that we can report just recent immigrants (in the country for one year or less). For Germany, we report all Polish immigrants. We smooth estimates in the US by taking a moving average over a three-year period $(t-1, t, t+1)$, in each year.

In Figure A3, we plot the evolution of the stock of Polish immigrants in these three countries as estimated from German, UK, and US microdata (solid line) and from PLFS data (scattered line).[27] The estimates from these independent datasets are reassuringly similar, showing very similar trends across data sources. We also compute the 95 percent confidence interval for the difference in the two data series. For Germany, this difference is only statistically significant for the first three years (note that estimates are very precise, because of the large sample size of the German administrative data); for the UK, the difference is statistically significant only in 2007.

[27] We use the IABS, UK LFS, and CPS data for Germany, the UK, and the US, respectively. See Section I for details on these datasets. Because the UK LFS and CPS contain information on years since migration, we can, in this case, focus on recent emigrants. Thus, in the figure, we plot the number of Poles who have been in the UK or the US for a period of less than two years. For the US data, however, the measurement of Polish immigrants is noisy because of small sample sizes, so in the figure, we smooth the graph using a three-year moving average.

Table A1. *Average annual number of observations by region*

	Emigrants		Total population	
	Observations	Weighted	Observations	Weighted
Lower Silesian	121.2	22,709	11,364.5	2,004,362
Kuyavian-Pomeranian	72.7	11,739	9,810.0	1,478,354
Lublin	148.5	22,045	10,658.8	1,559,638
Lubusz	58.4	6,589	6,994.9	751,529
Lódkie	55.4	10,419	11,543.5	2,016,047
Lesser Poland	214.0	38,033	12,524.0	2,173,142
Masovian	55.5	14,834	13,946.8	3,425,218
Opole	143.8	15,409	6,181.0	670,245
Subcarpathian	301.2	41,789	10,018.7	1,354,768
Podlaskie	149.6	17,523	6,574.8	767,856
Pomeranian	72.9	12,068	8,710.4	1,346,916
Silesian	69.1	16,546	14,521.8	3,205,886
Swietokrzyskie	112.4	13,959	8,035.3	946,403
Warmian-Masurian	74.7	9,764	7,819.3	976,397
Greater Poland	71.5	13,871	13,475.5	2,324,207
West Pomeranian	74.7	11,012	8,149.9	1,134,917
Mean	112.2	17,394	10,020.6	1,633,493

Source: PLFS.

Notes: We report the average number of annual observations, and the corresponding weighted figures, for emigrants and non-emigrants by region, over the years 1998–2007.

Table A2. *Stock of emigrants by destination country and year*

	Germany	US	UK	Ireland	Italy	Spain	France	Netherlands	Belgium	Sweden	Austria	Other
1994	32%	26%	2%	0%	7%	2%	3%	2%	2%	1%	4%	18%
1995	32%	26%	3%	0%	9%	3%	5%	1%	2%	0%	4%	15%
1996	30%	28%	3%	0%	8%	3%	5%	1%	2%	1%	4%	16%
1997	28%	32%	3%	0%	8%	3%	4%	1%	2%	0%	2%	16%
1998	27%	30%	4%	0%	11%	2%	4%	1%	2%	0%	3%	16%
1999	28%	28%	4%	0%	13%	3%	3%	1%	4%	0%	3%	13%
2000	36%	22%	6%	0%	7%	2%	4%	2%	3%	0%	4%	13%
2001	37%	21%	6%	1%	11%	2%	4%	3%	4%	1%	3%	8%
2002	35%	22%	6%	1%	12%	2%	3%	3%	4%	1%	2%	8%
2003	31%	19%	9%	1%	13%	3%	5%	5%	4%	1%	2%	7%
2004	27%	18%	14%	2%	13%	4%	5%	3%	2%	1%	3%	7%
2005	22%	13%	23%	6%	11%	4%	4%	3%	2%	2%	2%	8%
2006	18%	9%	32%	9%	7%	3%	3%	3%	2%	2%	2%	10%
2007	16%	6%	33%	12%	7%	3%	3%	5%	1%	1%	2%	10%
2008	15%	6%	33%	11%	6%	4%	3%	6%	2%	2%	2%	10%

Source: PLFS.

Notes: We report for each year the distribution of emigrants across destination countries.

The differences between the CPS and PLFS estimates for Polish immigrants are never statistically significant. Overall, therefore, these figures suggest that the emigration data we are using are quite accurate.

Appendix B: Instrumental Variables Construction and Estimation

In this appendix, we provide details about the way we construct our instrumental variables, and on how we measure the economic shocks to the main destination countries these variables are based on.

Instrumental Variables Estimation

We consider the four countries to which the majority of Polish emigrants migrated over the period under consideration: Germany, Ireland, the UK, and the US – countries where about 65 percent of all emigrants settled between 1998 and 2007. For each of these countries, we define a variable Z_c^t that captures the attractiveness of the respective destination country c for potential migrants. Each variable Z_t^c is expected to be correlated with the inflow of immigrants into country c but should not be correlated with any shock specific to a particular Polish region. Note that any possible correlations of Z_t^c with economic shocks that are common to all Polish regions are fully captured by the time dummies. Our exclusion restriction is that shocks to the destination countries have no region-specific consequences for Poland, apart from changing relative gains from migration.

In our preferred specification, we define Z_t^c as the annual growth rate of real wages at those parts of the wage distribution or in those sectors where Polish immigrants are most likely to be employed in each destination country c,[28] expressed in Polish zloty (for details, see the final subsection *Construction of Pull Factors* of this appendix). We then allow the effect of each Z_t^c to differ across different Polish regions r by interacting Z_t^c with regional dummies, and we define $Z_{rt}^c = Z_t^c \times R_r$. Finally, we account for the change in the relative role of economic shocks in different countries on migration decisions caused by the 2004 EU enlargement. We define two dummy variables EU_1 and EU_2 that identify the period in which Poland was not an EU member (up to and including 2003) and the years after Poland joined the EU (2004 onwards), respectively. We then interact Z_{rt}^c with EU_p $(p = 1, 2)$ and define $Z_{rpt}^c = Z_{rt}^c \times EU_p$.

This results in a vector $\mathbf{Z}$ of 120 instruments: that is, four destination countries $\times$ 16 regions $\times$ two time periods $=$ 128, of which eight (one region for each country for the years before and after 2004) must be set

[28] We use the growth rate of average wages below the 40th percentile for Germany, the UK, and the US, and the growth rate of wages in the construction and manufacturing sectors for Ireland.

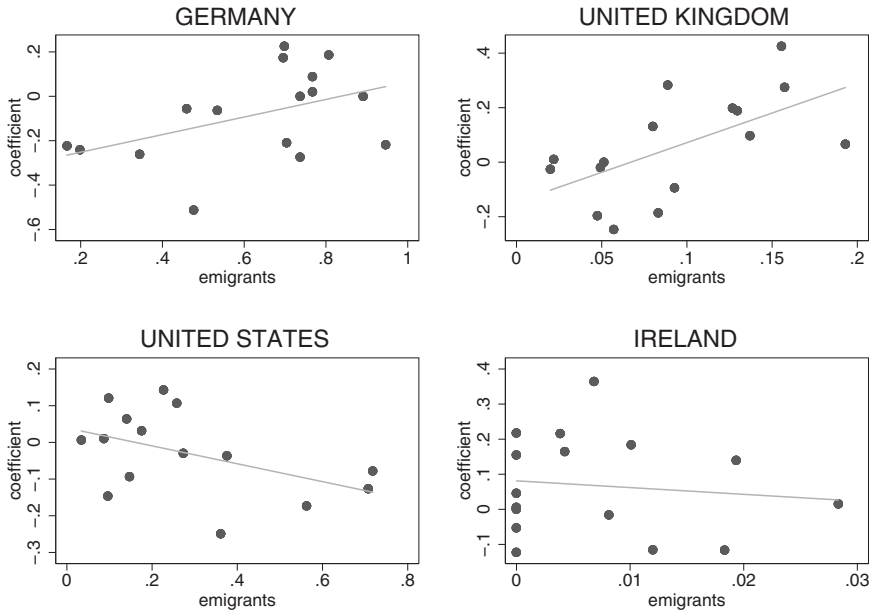

Fig. B1. Pre-2004 first-stage coefficients and emigrant share by destination country and region

Source: PLFS.

Notes: In each panel, we plot the first-stage coefficient pre-2004 for each region versus the mean share of emigrants in the same region. Each coefficient measures the effect of a wage shock in the destination country on emigration from each Polish region. The measure of emigrants we use is the average percentage of emigrants in the region to the destination country between 1998 and 2004 (excluded).

to zero for normalization. Our first-stage regression is thus

$$m_{rt} = \sum_{c\in(DE,IE,UK,US)} \sum_{r=1}^{R} \sum_{p=1}^{2} b_{crp} Z^{c}_{rpt} + X_{rt} g + \sum_{r=1}^{R} d_r R_r + \sum_{t=1}^{T} f_t \tau_t + \nu_{rt}. \quad \text{(B1)}$$

Each coefficient b_{crp} captures the effect that a shock to destination country c has on the emigration rate in region r before ($p = 1$) or after ($p = 2$) Poland joined the EU, net of time-invariant regional characteristics R_r, nationwide time-variant shocks τ_t, and other exogenous factors X_{rt}. We expect shocks to country c to have a higher impact on emigration (i.e., b_{crp} to be larger) in regions in which a larger fraction of total emigration is directed to that destination country. In Figures B1 and B2, we plot

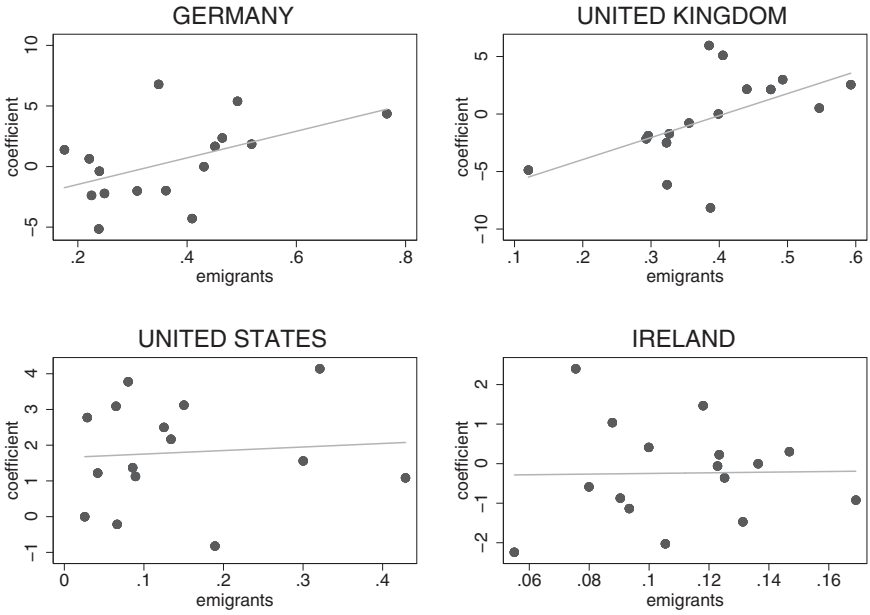

Fig. B2. Post-2004 first-stage coefficients and emigrant share by destination country and region

Source: PLFS.

Notes: In each panel, we plot the first-stage coefficient post-2004 for each region versus the mean share of emigrants in the same region. Each coefficient measures the effect of a wage shock in the destination country on emigration from each Polish region. The measure of emigrants we use is the average percentage of emigrants in the region to the destination country between 2004 (included) and 2007.

the estimated coefficients b_{crp} versus the fraction of the mean number of emigrants over the mean population in period p from each region r in each destination country c for the years before and after 2004, respectively. As the figures show, the coefficients that weigh shocks in each country to different regions are positively correlated with the regional fraction of emigrants to that destination country except in the case of the US and Ireland[29] for the years before 2004. This correlation reassures us that our coefficients are picking up actual effects of destination country shocks on regional emigration.

[29] However, it should be noted that the share of Polish emigrants in Ireland before 2004 was extremely low, and migration to the US also declined sharply over this period.

Alternative Definition of Instruments

One concern with the IV strategy described above is that the number of instruments (120) is very high, and close to the number of observations (160). In this case, the IV estimate might be biased toward the OLS estimate. For this reason, we check the robustness of our IV estimates to alternative definitions of instruments where we use different strategies to reduce the dimensionality. First, we select only the variables that are individually statistically significant. We estimate equation (B1) and select only those variables for which the coefficient is statistically significant at the 1 percent level. This procedure reduces the number of instruments to 44. Second, we use two different model selection algorithms: backward elimination and forward selection. In the backward-elimination model, we start from the full set of instruments, and drop at every step the variable that is least significant, provided that the p-value of a t-test for the null of a zero coefficient is higher than 10 percent. The procedure stops when estimated coefficients for all the included variables are statistically significant at 10 percent. This algorithm leads to the selection of 77 instruments. Similarly, in the forward-selection model, we start from the model estimated with just a constant term, and then add, alternatively, each of the other candidate instruments. We then select the variable with the highest statistical significance, add it to the model, and then iterate the procedure, selecting at each step one additional variable until no additional variable is statistically significant at 10 percent. This procedure reduces the number of instruments to 23.

Our third strategy is to model the differential effects of destination country shocks on Polish regions with pre-assigned continuous regional weights (ω_r^c) and to use $\omega_r^c \times Z_t^c \times EU_p$ as instruments, rather than relying on estimated region–country pair specific weights. Doing so reduces the dimensionality of our instruments vector $\mathbf{Z}$ to 8. We use two alternative weights. First, we use the inverse of the distance between each region's capital and the capital of the destination country c. The assumption in this case is that migration costs increase with distance, so the shocks to destination countries should have a higher pull effect on regions that are closer. Second, because a number of papers have illustrated the importance of migration networks on migration decisions (e.g., Bartel, 1989; Munshi, 2003), we expect shocks from country c to have a stronger pull effect on emigration from regions in which a higher share of individuals had previously emigrated to that specific country. Because we have no reliable data on historical regional emigration to different destination countries, we cannot measure the historical strength of regional migration networks. However, we can measure the strength of regional migration networks to destination country c using the mean share of emigrants from region r to country c

over the 1998–2007 period. In constructing this variable, we reduce possible feedback by excluding, in every year t, the share of emigrants in year $t-1$, t, and $t+1$. This means that, in practice, for every year t_0, we define weights $\omega^c_{rt_0}$,

$$\omega^c_{rt_0} = \frac{1}{N_{t_0}} \sum_{t<t_0-1, t>t_0+1} emigrants^{\mathrm{c}}_{\mathrm{rt}}/emigrants_{\mathrm{rt}},$$

where N_{t_0} is the number of years over which the mean is computed in year t_0.

Finally, we check the robustness of our results to the use of alternative variables as pull factors Z^c_t. As explained above, in our baseline results, we define Z^c_t as the growth rate of average wages below the 40th percentile, expressed in zloty, in destination country c. Here, we experiment with one alternative: the deviation of the national per capita GDP growth in year t for country c, GDP^c_t, from the OECD mean GDP growth,[30] GDP^{OECD}_t; see McKenzie *et al.* (2014) for evidence on the role of the GDP growth in destination countries on migration choices. Because the resulting variable captures the relative economic performance of country c relative to other OECD countries, we expect it to measure the nation's relative attractiveness for potential migrants.

Construction of Pull Factors

In this subsection, we provide details on our choice of pull factors for the construction of our main instrumental variables.

In our preferred specification, we define Z^j_t as the annual growth rate of real wages below the 40th percentile in country j (US, UK, or Germany) and wages in the construction and manufacturing sector for Ireland, expressed in zloty. We choose the growth rate of mean wages below the 40th percentile to measure the attractiveness of destination countries for Polish emigrants because these fall into the lower part of the wage distribution in host countries, especially in the first years after migration. This fact is demonstrated in Figure B3, which uses IABS, CPS, and UK LFS data, respectively, to plot the position of Polish immigrants in the wage distribution for Germany, the US, and the UK (for all three countries, Poles are in the lower part of the distribution for natives). For Ireland, for which we have no microdata but only aggregate data by industry, the 2006 Irish Census indicates that over half of Polish male immigrants are working in construction

[30] In both cases, the GDP is in US constant dollars. Source: OECD *Statistical Extracts* (http://stats.oecd.org/Index.aspx).

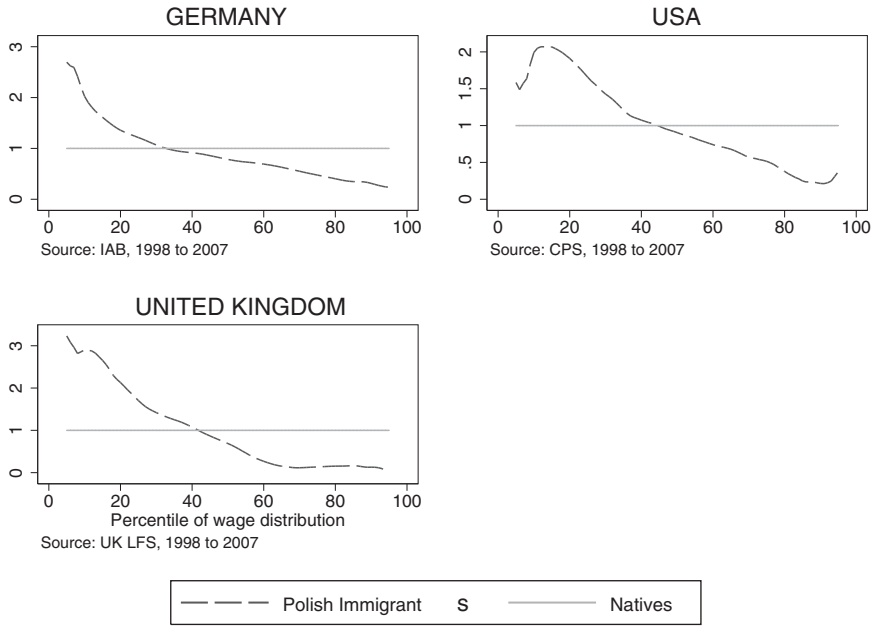

Fig. B3. Polish emigrants in the destination countries' wage distribution
Source: PLFS.
Notes: The graphs report the relative distribution of Polish immigrant wages in the wage distribution of natives in Germany, the US, and the UK, for years 1998–2007 pooled, using data from the destination countries. For the UK and the US, the figures refer to recent immigrants (less than two years in the country); for Germany, the figure refers to all Polish emigrants.

and manufacturing (Irish Central Statistical Office, 2008).[31] Moreover, we calculate wages in zloty, which accounts for fluctuations in the exchange rate of the US dollar, the British pound, and the euro with regards to the Polish currency. This latter is important because a large portion of the earnings is likely to be spent in Poland either through reallocation to families in the form of remittances or through the temporariness of migration, which means that savings are later spent at home (e.g., Dustmann, 1997). As Table B1 shows, such exchange rate fluctuations were sizeable during the years under consideration and contributed substantially to changes in the earnings differential of Polish workers in Poland and abroad, in terms of their purchasing power in Poland.

[31] Defining Z_t^j as the growth rate of mean real wages, or of real wages below the median, results in weaker instruments but leads to qualitatively similar results.

Table B1. *Zloty exchange rates with respect to the US dollar, the British pound, and the euro*

	PNL/$	PNL/euro	PNL/£
1998	3.48	3.90	5.76
1999	3.97	4.23	6.42
2000	4.35	4.00	6.58
2001	4.09	3.66	5.89
2002	4.08	3.84	6.11
2003	3.89	4.39	6.35
2004	3.66	4.54	6.70
2005	3.24	4.02	5.88
2006	3.10	3.89	5.71
2007	2.77	3.79	5.54

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

References

Altonji, J. G. and Card, D. (1991), The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives, in J. M. Abowd and R. B. Freeman (eds.), *Immigration, Trade and the Labor Market*, University Chicago Press, Chicago, IL, 201–234.

Amuedo-Dorantes, C. and Pozo, S. (2006), Migration, Remittances, and Male and Female Employment Patterns, *American Economic Review 96* (2), 222–226.

Angrist, J. D. and Kugler, A. D. (2003), Protective or Counter-Productive? Labour Market Institutions and the Effect of Immigration on EU Natives, *Economic Journal 113* (488), F302–F331.

Antman, F. M. (2013), The Impact of Migration on Family Left Behind, in A. F. Constant and K. F. Zimmermann (eds.), *International Handbook on the Economics of Migration*, Edward Elgar, Northampton, MA, 293–308.

Aydemir, A. and Borjas, G. J. (2007), Cross-Country Variation in the Impact of International Migration: Canada, Mexico, and the United States, *Journal of the European Economic Association 5*, 663–708.

Bartel, A. (1989), Where Do the New U.S. Immigrants Live?, *Journal of Labor Economics 7*, 371–391.

Bianchi, M., Buonanno, P., and Pinotti, P. (2012), Do Immigrants Cause Crime?, *Journal of the European Economic Association 10*, 1318–1347.

Borjas, G. J. (2003), The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market, *Quarterly Journal of Economics 118*, 1335–1374.

Borjas, G. J., Freeman, R. B., and Katz, L. F. (1996), Searching for the Effect of Immigration on the Labor Market, *American Economic Review 86* (2), 246–251.

Borjas, G. J., Freeman, R. B., and Katz, L. F. (1997), How Much Do Immigration and Trade Affect Labor Market Outcomes?, *Brookings Papers on Economic Activity 1997* (1), 1–90.

Bratsberg, B. and Raaum, O. (2012), Immigration and Wages: Evidence from Construction, *Economic Journal 122* (565), F1177–F1205.

Bratsberg, B., Raaum, O., Røed, M., and Schøneb, P. (2014), Immigration Wage Effects by Origin, *Scandinavian Journal of Economics 116*, 356–393.

Card, D. (2001), Immigrant Inflows, Native Outflows and the Local Labor Market Impacts of Immigration, *Journal of Labor Economics 19*, 22–64.

Card, D. and Lewis, E. G. (2007), The Diffusion of Mexican Immigrants during the 1990s: Explanations and Impacts, in G. J. Borjas (ed.), *Mexican Immigration*, University of Chicago Press (for NBER), Chicago, IL, 193–227.
Card, D. and DiNardo, J. (2000), Do Immigrant Inflows Lead to Native Outflows?, *American Economic Review 90* (2), 360–367.
Chiquiar, D. and Hanson, G. H. (2005), International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States, *Journal of Political Economy 113*, 239–281.
Cortes, P. (2008), The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data, *Journal of Political Economy 116*, 381–422.
D'Amuri, F., Ottaviano, G. I. P., and Peri, G. (2010), The Labor Market Impact of Immigration in Western Germany in the 1990s, *European Economic Review 54*, 550–570.
Docquier, F., Ozden, C., and Peri, G. (2014), The Labor Market Impact of Immigration and Emigration in OECD Countries, *Economic Journal 124* (579), F1106–F1145.
Drinkwater, S., Eade, J., and Garapich, M. (2009), Poles Apart? EU Enlargement and the Labour Market Outcomes of Immigrants in the United Kingdom, *International Migration 47*, 161–190.
Dustmann, C. (1997), Return Migration, Uncertainty and Precautionary Savings, *Journal of Development Economics 52*, 295–316.
Dustmann, C., Fabbri, F., and Preston, I. (2005), The Impact of Immigration on the UK Labour Market, *Economic Journal 115* (507), F324–F341.
Dustmann, C., Frattini, T., and Preston, I. (2013), The Effect of Immigration along the Distribution of Wages, *Review of Economics Studies 80*, 145–173.
Elsner, B. (2013a), Does Emigration Benefit the Stayers? Evidence from EU Enlargement, *Journal of Population Economics 26*, 531–553.
Elsner, B. (2013b), Emigration and Wages: The EU Enlargement Experiment, *Journal of International Economics 91*, 154–163.
Fernández-Huertas Moraga, J. (2011), New Evidence on Emigrant Selection, *Review of Economics and Statistics 93*, 72–96.
Frattini, T. (2010), Immigration and Prices in the UK, Unpublished manuscript, University of Milan.
Glitz, A. (2012), The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany, *Journal of Labor Economics 30*, 175–213.
Hanson, G. (2007), Emigration, Labor Supply and Earnings in Mexico, in G. J. Borjas (ed.), *Mexican Immigration*, University of Chicago Press (for NBER), Chicago, IL, 289–328.
Hatton, T. J. and Williamson, J. G. (2006), International Migration in the Long-Run: Positive Selection, Negative Selection and Policy, in F. Foders and R. J. Langhammer (eds.), *Labour Mobility and the World Economy*, Springer, Berlin, 1–31.
Irish Central Statistical Office (2008), Census 2006: Non-Irish Nationals Living in Ireland. Profiles of Nationalities – UK, Polish, Lithuanian, Nigerian, Latvian, available online at http://www.cso.ie/en/media/csoie/census/documents/PROFILES%20OF%20NATIONALITIES%201-5.pdf
Jaeger, D. A. (2007), Skill Differences and the Effect of Immigrants on the Wages of Natives, Mimeo, College of William and Mary, Williamsburg, VA.
Kaczmarczyk, P. and Okólski, M. (2008), Demographic and Labour-Market Impacts of Migration on Poland, *Oxford Review of Economic Policy 24*, 599–624.
Kaczmarczyk, P., Mioduszewska, M., and Zylicz, A. (2009), Impact of the Post-Accession Migration on the Polish Labour Market, in M. Kahanec and K. F. Zimmermann (eds.), *EU Enlargement and the Labor Markets: What Do We Know?*, Springer, Berlin, 219–253.
Kraetke, S. (1996), Where East meets West – The German–Polish Border Region in Transformation, *European Planning Studies 4*, 647–669.

Kraetke, S. (1999), Regional Integration or Fragmentation? The German–Polish Border Region in a New Europe, *Regional Studies 33*, 631–641.
Lewis, E. G. (2011), Immigration, Skill Mix, and Capital–Skill Complementarity, *Quarterly Journal of Economics 126*, 1029–1069.
McKenzie, D. and Rapoport, H. (2007), Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico, *Journal of Development Economics 84*, 1–24.
McKenzie, D., Theoharides, C., and Yang, D. (2014), Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks, *American Economic Journal: Applied Economics 6*, 49–75.
Manacorda, M., Manning, A., and Wadsworth, J. (2012), The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain, *Journal of the European Economic Association 10*, 120–151.
Mishra, P. (2007), Emigration and Wages in Source Countries: Evidence from Mexico, *Journal of Development Economics 82*, 180–199.
Munshi, K. (2003), Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market, *Quarterly Journal of Economics 118*, 549–599.
Ottaviano, G. I. P. and Peri, G. (2012), Rethinking the Effects of Immigration on Wages, *Journal of the European Economic Association 10*, 152–197.
Pedersen, P. J., Pytlikova, M., and Smith, N. (2008), Selection and Network Effects – Migration Flows into OECD Countries 1990–2000, *European Economic Review 52*, 1160–1186.
Stock, J. H. and Yogo, M. (2005), Testing for Weak Instruments in Linear IV Regression, in D. W. K. Andrews (ed.), *Identification and Inference for Econometric Models*, Cambridge University Press, New York, 80–108.