The Impact of Immigration along the Distribution of Native Wages:

Evidence from Germany in the 2000s

Ararat Gocmen

Supervisor: Christian Dustmann

Dissertation submitted in part-fulfilment of the
MSc in Economics, University College London

September 2020

Declaration

I, Ararat Gocmen, hereby declare that the work presented in this dissertation is my own original work. Where information has been derived from other sources, I confirm that this has been clearly and fully identified and acknowledged. No part of this dissertation contains material previously submitted to the examiners of this or any other university, or any material previously submitted for any other assessment.

Name: Ararat Gocmen

Date: September 18, 2020

Classification

This piece of research is primarily:

- ☒ an empirical/econometric study.
- ☐ the examination of a theoretical problem.
- ☐ a critical analysis of a policy issue.
- ☐ an analytical survey of empirical and/or theoretical literature.

Dissertation Library

I give permission for a PDF electronic copy of this dissertation to be made available to future students by way of the Dissertation Library, should the Department of Economics choose to use it in this way:

Yes / No

Abstract

Given inelastic labor supplies for different skill groups and an elastic capital supply, theory predicts that immigration decreases native wages for groups in which immigrants are overrepresented compared to natives, and increases them for those in which immigrants are underrepresented. I provide empirical evidence for this theory in the case of the immigration waves into Germany that began in 2011, which by 2017 raised the immigrant share of the full-time labor force in the former states of West Germany (excluding Berlin) by almost 3 percentage points. I first demonstrate that immigrants are more likely than natives to work in lower-wage occupations, such that immigrants are overrepresented below the first quartile of the native wage distribution and underrepresented above it. Immigrant downgrading amplifies these disparities, as immigrants earn lower wages compared even to natives of similar qualifications. I then estimate the impact of immigration on different percentiles of the native wage distribution. My findings are consistent with both theory and the evidence of downgrading: the impact is negative below the first quartile and positive above it. Causal identification is achieved by differencing across time to eliminate regional fixed effects, and by instrumenting the change in the immigrant-to-native ratio by its lagged level to account for various sources of bias. Based on this identification strategy, the recent immigration waves increased inequality in the bottom half of the native wage distribution in the average West German state, widening the wage differential between native workers at the median and 5th percentile by about 6%.

Keywords: Immigration, Wage inequality, Germany
JEL codes: J21, J31, J61

Acknowledgements

I would first and foremost like to thank Professor Christian Dustmann, without whose guidance and published research this dissertation would have been impossible to write. He also connected me with his student Carl Gergs, whose data-cleaning tips and project-management advice were invaluable.

I would also like to thank Clara von Bismarck-Osten, Maren Holthe Hedne, Allen Shen, and Emily Clayton for their company and reassurance this summer, as well as Vishan Nigam, René Chalom, Naman Jain, and Andrew Tynes for their help throughout my first foray into economics research.

Word count: 9,896

Contents

List of Figures

List of Tables

1 Introduction

Immigration to Germany has surged in the past decade (Higgins and Klitgaard 2019). A first immigration wave started in 2011, as workers from other countries in Europe began moving to Germany due to its relatively stronger economic performance since the dawn of the global financial crisis. The introduction in May 2011 of free movement for workers from Poland and the seven other new member states that joined the European Union in 2004, and in January 2014 for workers from Bulgaria and Romania which joined in 2007, explains why this first wave has primarily been driven by Central and Eastern European immigrants (Bertoli, Brücker, and Fernández-Huertas Moraga 2013). The German government's response to the refugee crisis in Syria, Afghanistan, and other war-torn countries in the Middle East, Asia, and Africa launched a second immigration wave in 2015 (*The Economist* 2020).

In this paper, I analyze the impact of these immigration waves along the distribution of native German workers' wages. As new immigrants have entered labor markets, they have changed the relative supply of and demand for different skills in Germany. Some native workers have suffered decreases in their wages due to the increased competition with immigrants, while others have even enjoyed increases as a result of the second-order effects of this labor supply shock. Immigration has therefore changed the structure of native wage inequality. An evaluation of this differential impact of immigration on native wages can identify the winners and losers of the past decade of open immigration policies in Germany, potentially informing the cost-benefit analysis behind future decisions to further loosen or begin restricting immigration. It can also shed light on reasons for the rise of anti-immigration sentiments among German voters and their support for political parties like Alternative for Germany that give voice to such sentiments.

Following the approach in Dustmann, Frattini, and Preston (2013), I contribute to the international evidence supporting the authors' theory of the differential impact of immigration on native wages. Derived and explained in Section 2, this theory predicts that the impact of immigration on the wages of native workers in a particular skill group is a function of the density of immigrants relative to natives in that group. Native wages decrease for skill groups in which immigrants are overrepresented compared to natives. However, they can increase for groups in which immigrants are underrepresented if the capital supply is sufficiently elastic, since the capital inflows induced by immigration can increase native marginal productivity enough to offset the downward pressure from increased labor-market competition. This section also motivates an empirical specification based on this theory and justifies an identification strategy allowing for a causal interpretation of its estimates.

Section 3 surveys the existing research on the impact of immigration on native wages, based primarily on the comprehensive review of the literature in Dustmann, Schönberg, and Stuhler (2016) but also considering more recent publications. My econometric approach is compared to others detailed in this review, and the estimates from studies using the pure spatial approach to whose category mine belongs are discussed.

Section 4 describes the German data and my focus on only the full-time labor force in the former states of West Germany (excluding Berlin). It then provides descriptive evidence of immigrant overrepresentation in lower-wage occupations, partially attributing this overrepresentation to their downgrading into ones below their qualifications. Based on relative density estimates, immigrants are overrepresented below the first quartile of the native wage distribution and underrepresented above it, and more so than they would be if they earned the same returns to their observable characteristics as natives.

Section 5 finally reports the estimates of the main empirical specification. As predicted by theory, the estimated impact of immigration along the distribution of native wages from 2011 to 2017 mirrors the estimated relative density of immigrants, with a negative impact below the first quartile and a positive one above it. While these estimates are not statistically significant at all percentiles of the native wage distribution, they are at the bottom and middle of it. A 1 percentage-point increase in the immigrant-to-native ratio decreases the 5th percentile of native wages by slightly more than 1% while increasing the median by about the same amount, increasing inequality in the bottom half of the distribution. The magnitudes of these estimates are large compared to most in the existing literature, but the distribution of their directions—negative for the poorest natives and positive for middling ones—is consistent with published results.

Given the 3 percentage-point increase in the immigrant-to-native ratio since 2010, the recent immigration waves have increased the 50-5 wage differential among natives by about 6% in the average West German state, a conclusion whose provocative implications are considered in Section 6.

2 Theory and econometric approach

2.1 The model

In the basic theoretical model developed in Dustmann, Frattini, and Preston (2013), wages are determined in partial equilibrium by identical perfectly competitive firms that face inelastic supplies of different types of labor, and that are equipped with the two-level constant elasticity of substitution production technology characterized by the following equations:

$$Y = [\beta H^s + (1-\beta)K^s]^{\frac{1}{s}} \tag{1}$$

$$H = \left(\sum_{i=1}^{n} \alpha_i l_i^{\sigma}\right)^{\frac{1}{\sigma}} \tag{2}$$

Equation (1) describes the production of a single type of output Y using human capital H and (physical) capital K as inputs, and equation (2) the production of human capital using n different types of labor l_i as inputs. The elasticity of substitution between human capital and capital is determined by the parameter $s < 1$, and their relative productivities by the parameter $0 < \beta < 1$. Similarly, the elasticity of substitution between the different types of labor is determined by the parameter $\sigma < 1$, and their relative productivities by the parameters $0 < \alpha_i < 1$ that satisfy the condition $\sum_{i=1}^{n} \alpha_i = 1$.

For each type of labor, native labor l_i^N and immigrant labor l_i^I are perfectly substitutable, such that the supply of a specific type is $l_i = l_i^N + l_i^I$. The native labor supply is $L^N = \sum_{i=1}^{n} l_i^N$, and each type's fraction of it $\pi_i^N = \frac{l_i^N}{L^N}$. Similarly, the immigrant labor supply is $L^I = \sum_{i=1}^{n} l_i^I$, and each type's fraction of it $\pi_i^I = \frac{l_i^I}{L^I}$. The total immigrant-to-native ratio is $m = \frac{L^I}{L^N}$. The supply of each type of labor can then be rewritten as $l_i = L^N(\pi_i^N + m\pi_i^I)$.

Given perfect competition, each factor of production earns its marginal productivity. Derived in Appendix A.1, the first-order conditions of the representative profit-maximizing firm, which sells output at the normalized price 1 while hiring each type of labor at the wage w_i and capital at the rate ρ, are:

$$\ln w_i = \ln(\beta\alpha_i) + (\sigma - 1)\ln\left(\pi_i^N + m\pi_i^I\right) + (1-\sigma)\ln\left(\frac{H}{L^N}\right)$$
$$+ \left(\frac{1}{s} - 1\right)\ln\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^s\right] \tag{3}$$

$$\ln \rho = \ln(1-\beta) + (s-1)\ln\left(\frac{K}{H}\right)$$
$$+ \left(\frac{1}{s} - 1\right)\ln\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^s\right] \tag{4}$$

Equations (3) and (4) suggest multiple channels by which the labor supply shock of immigration affects each type's wage by changing its marginal productivity. The second term on the right-hand side of equation (3) captures the negative effect of increased competition within each type due to an increase in the immigrant population. Offsetting this negative effect, the third term captures the positive effect from the complementarity of different types: the employment of more immigrants of all types increases human capital through equation (2) and thereby the marginal productivity of each type. If this increase in human capital also increases the rate of return on capital through equation (4), thereby inducing an increase in the supply of capital, there is an additional positive effect captured in the fourth term of equation (3).

The impact of the immigrant-to-native ratio m on each type's wage can be stated more precisely. Derived in Appendix A.2, it is:

$$\frac{\partial \ln w_i}{\partial m} = \gamma_i = (\sigma - 1)\left(\frac{\pi_i^I}{\pi_i^N + m\pi_i^I} - \phi\sum_{i=1}^{n}\omega_i\frac{\pi_i^I}{\pi_i^N + m\pi_i^I}\right) \tag{5}$$

where ϕ reflects the capital supply elasticity, equaling 1 if the capital supply is perfectly elastic and less than 1 if not; and the weight ω_i reflects each type's contribution to human capital production. If there are initially no immigrants such that $m = 0$, equation (5) simplifies to:

$$\frac{\partial \ln w_i^0}{\partial m} = \gamma_i^0 = (\sigma - 1)\left(\frac{\pi_i^I}{\pi_i^N} - \phi\sum_{i=1}^{n}\omega_i^0\frac{\pi_i^I}{\pi_i^N}\right) \tag{6}$$

where $w_i^0 = w_0|_{m=0}$ is each type's wage and $\omega_i^0 = \omega_i|_{m=0}$ reflects their contribution to human capital production before immigration.

According to equation (6), the impact of immigration on native wages of a given type is a function of the density of immigrants relative to natives of that type, and of the capital supply elasticity. Recall that $\sigma < 1$. If the capital supply is perfectly elastic (i.e. $\phi = 1$), immigration then leads to a decrease in a given type's wages if immigrants are overrepresented in that type compared to natives, and an increase if they are underrepresented. Intuitively, the downward pressure due to increased labor-market competition that immigration puts on native wages of a given type is increasing in the relative density of immigrants of that type. Though there are upward pressures from the marginal productivity gains due to complementarity between types and the capital inflows induced by immigration, these offsetting effects are enjoyed more by types facing less competition from immigrants. The driving force behind the differential impact of immigration along the distribution of types is thus the relative density of immigrants across types.

If instead the capital supply is relatively inelastic (i.e. $\phi < 1$), then immigration can lead to a decrease in the wages of all types. Intuitively, if immigration fails to induce capital inflows that increase native marginal productivity, then the negative effect of increased competition is the dominant one of immigration, and even native types underrepresented in the immigrant population experience wage losses.

2.2 The empirical specification

The different types of labor in this theoretical model represent the different skill groups of workers discussed in Section 1. It would be intuitive to define these skill groups according to the observable characteristics of workers that proxy for their skills, for example, their education or experience levels. Dustmann, Frattini, and Preston (2013) instead define skill groups based on workers' ranks in the distribution of native wages. This definition identifies the same groups of native workers before and after immigration—and which is therefore valid for the purpose of evaluating the impact of immigration on the wages of these groups—as long as immigration does not change their ranks in the native wage distribution. Derived in Appendix A.3, a necessary and sufficient condition for rank insensitivity to immigration between two different skill groups i and j that satisfy the conditions $w_i^0 > w_j^0$ and $w_i > w_j$ is:

$$\frac{\frac{\pi_i^I}{\pi_i^N} - \frac{\pi_j^I}{\pi_j^N}}{\ln\left(\frac{w_i}{w_j}\right)} \geq \frac{1}{(\sigma-1)m} \tag{7}$$

Condition (7) is empirically verifiable given data on wages and the relative density of immigrants at different percentiles of the native wage distribution, as well as estimates of σ and m. If this condition holds, then different labor types can be defined by these percentiles, and the skill group indicators i and j represent the different percentiles considered in the main empirical specification estimated in this paper.

This specification is motivated theoretically by a first-order Taylor approximation of equation (3) around $m = 0$, derived in Appendix A.4:

$$\ln w_i \approx \ln(\beta\alpha_i) + (\sigma-1)\ln\left(\pi_i^N\right) + \frac{1-\sigma}{\sigma}\ln\left(\sum_{i=1}^{n}\alpha_i(\pi_i^N)^\sigma\right)$$
$$+ G(\rho_0) + (\sigma-1)\zeta_i^0 m \tag{8}$$

where ρ_0 is the rate of return on capital, $G(\rho_0)$ is a function evaluated at it, and $\zeta_i^0 = \frac{\pi_i^I}{\pi_i^N} - \phi\sum_{i=1}^{n}\omega_i^0\frac{\pi_i^I}{\pi_i^N}$ is the relative density of immigrants of a given type, all in the case when there are initially no immigrants. If skill groups are defined by their percentiles in the native wage distribution in a given region at a given time (i.e. $i = prt$), then variation across regions and times can be used to translate expression (8) into the following fixed-effects model:

$$\ln w_{prt} = a_{pr} + b_{pt} + \gamma_{prt}^0 m_{rt} + c_p x_{rt} + \epsilon_{prt} \tag{9}$$

where w_{prt} is the real wage at the pth percentile of the native wage distribution in region r at time t, a_{pr} are regional fixed effects for each percentile constant across time, b_{pt} are time fixed effects for each percentile constant across regions, $\gamma_{prt}^0 = (\sigma-1)\zeta_{prt}^0$ is the impact of immigration on native wages when there are initially no immigrants, x_{rt} includes controls for the demographics of the native and immigrant labor forces, and ϵ_{prt} is a residual term. If capital is perfectly mobile such that the rate of return on capital is constant across regions, an assumption whose validity is considered in Section 5.1, then the regional fixed effects identify $a_{pr} = (\sigma-1)\ln\left(\pi_i^N\right) + \frac{1-\sigma}{\sigma}\ln\left(\sum_{i=1}^{n}\alpha_i(\pi_i^N)^\sigma\right)$, whereas the time fixed effects identify $b_{pt} = \ln(\beta\alpha_i) + G(\rho_0)$ also assuming constant labor productivity across regions.

If the relative density of immigrants at each percentile is constant across regions and time, i.e. $\zeta^0_{prt} = \zeta^0_p$, then the impact of immigration is also constant across regions and time, i.e. $\gamma^0_{prt} = \gamma^0_p$. If not, then replacing γ^0_{prt} with γ^0_p in equation (9) identifies the average impact of immigration at each percentile across regions and time:

$$\ln w_{prt} = a_{pr} + b_{pt} + \gamma^0_p m_{rt} + c_p x_{rt} + \epsilon_{prt} \tag{10}$$

2.3 Structural interpretation

If there are immigrants already present at the time of a new immigration wave, then a comparison of equations (5) and (6) shows that the impact of the inflow on native wages for a given type is in the same direction but of a smaller magnitude than in the case of no initial immigrants. Mathematically, the ratio $\frac{\pi^I_i}{\pi^N_i + m\pi^I_i}$ is decreasing in the immigrant-to-native ratio m for all types, and for small m the weights would not differ much (i.e. $\omega_i \approx \omega^0_i$). If the relative density of immigrants and the impact of immigration when immigrants are already present are respectively defined as $\zeta_i = \frac{\pi^I_i}{\pi^N_i + m\pi^I_i} - \phi \sum_{i=1}^n \omega_i \frac{\pi^I_i}{\pi^N_i + m\pi^I_i}$ and $\gamma_i = (\sigma - 1)\zeta_i$, then for $m > 0$:

$$\text{sign}(\zeta_i) = \text{sign}(\zeta^0_i) \text{ and } |\zeta_i| < |\zeta^0_i| \tag{11}$$

Next combine equations (5)-(6) and statement (11):

$$\text{sign}(\gamma_i) = \text{sign}(\gamma^0_i) \text{ and } |\gamma_i| < |\gamma^0_i| \tag{12}$$

According to statement (12), the magnitude of the impact of immigration on native wages is largest when there are initially no immigrants. Intuitively, in this case, newly arriving immigrants compete only with natives and put the maximal downward pressure on their wages, while the upward pressures due to complementarity across labor types and capital inflows are enjoyed only by natives. If instead there are already immigrants present, the new arrivals compete not only with natives but also the existing immigrants. Since the latter also partake in the resulting benefits of immigration, the effect of immigration on native wages are moderated in both directions.

This observation is important because the empirical specification in equation (10) is theoretically motivated by the assumption that there are no immigrants at the time of immigration. This assumption is arguably valid in the application considered in Dustmann, Frattini, and Preston (2013), since they use a definition of immigrant that distinguishes between immigrants who have only recently arrived to the country in question and the rest of the population, such that the assumption of $m = 0$ is reasonable.

In contrast, as detailed in Section 4.1, the German data that I use allows for only a broader definition of immigrant based on country of nationality. As a result, not only recent arrivals but also any non-natives who may have already been in Germany for an extended period of time are captured in and therefore inflate m, making the assumption of $m = 0$ less reasonable. The specification that I estimate is then not actually equation (10) but rather:

$$\ln w_{prt} = a_{pr} + b_{pt} + \gamma_p m_{rt} + c_p x_{rt} + \epsilon_{prt} \tag{13}$$

Again according to statement (12), the parameter γ_p identifies the direction and the lower bound of the magnitude of the true average impact of immigration on a percentile of the native wage distribution, were immigrants defined not only by their nationality but also by their duration of residence. Given an estimate $\widehat{\gamma}_p$ of the impact of immigration, theory would thus predict that the true impact is of the same direction but of greater magnitude.

Hereafter abstracting from this problem, I interpret the estimates of ζ_p and γ_p reported in Sections 4.3 and 5.1, respectively, as if they were of ζ_p^0 and γ_p^0—that is, as if there were initially no immigrants. Given these estimates at various percentiles, as well as the assumption of a perfectly elastic capital supply (i.e. $\phi = 1$), $(\sigma - 1)$ can be estimated by an ordinary-least squares (OLS) regression of $\widehat{\gamma}_p$ on $\widehat{\zeta}_p$ and a constant. The elasticity of substitution between different types of labor could then be estimated as $\frac{1}{1-\widehat{\sigma}}$.

2.4 Causal identification

Before estimating equation (13), I first take differences across times t and $t-1$ to eliminate regional fixed effects:

$$\Delta \ln w_{prt} = \Delta b_{pt} + \gamma_p \Delta m_{rt} + c_p \Delta x_{rt} + \Delta \epsilon_{prt} \tag{14}$$

The OLS estimates of γ_p from equation (14) are still likely to suffer from two sources of bias: measurement error due to small sample size, and immigrant selection into regions with higher wage growth.

On the one hand, if the number of immigrants in the sample used to calculate the immigrant-to-native ratio is small for any region at a given time, this ratio may be mismeasured, with differencing only exacerbating the problem of measurement error. The result is a downward attenuation bias in $\widehat{\gamma}_p^{OLS}$.

On the other hand, there is the problem of reverse causality: immigrants are likely to move to regions where wage growth is high, such that the OLS identifying assumption $\mathbb{E}(\Delta m_{rt}\Delta\epsilon_{prt}) = 0$ is violated. If instead $\mathbb{E}(\Delta m_{rt}\Delta\epsilon_{prt}) > 0$, then $\widehat{\gamma}_p^{OLS}$ also suffers from an upward bias.

To address both these sources of bias, I follow Dustmann, Frattini, and Preston (2013) in using an instrument variable (IV) for the *change* in the immigrant-to-native ratio Δm_{rt}: the *level* of this ratio $m_{rt-\tau}$ at a sufficient enough of lag τ to ensure exogeneity. This instrument is relevant if new immigrants are likely to move to regions with higher concentrations of existing immigrants, a phenomenon which occurs because new arrivals often have family, other social, or even economic ties to local immigrant communities.

The instrument $m_{rt-\tau}$ is exogenous to the measurement error in $\Delta\epsilon_{prt}$ if the measurement error in the immigrant-to-native ratio is serially uncorrelated. If measurement error in this ratio is a concern only due to small regional sample sizes rather than a systematic source of inaccuracy in labeling immigrants, then such serial correlation is unlikely.

The instrument also deals with the endogenous selection of immigrants into high-growth regions if shocks to wages are not too persistent, specifically, if $\mathbb{E}(\epsilon_{prt-\tau}\Delta\epsilon_{prt}) = 0$. Since $\Delta\epsilon_{prt} = \epsilon_{prt} - \epsilon_{prt-1}$, the ratio must be lagged by at least more than one period for this condition to hold. After choosing $\tau > 1$ and estimating equation (14) using an IV regression, the reasonability of this exogeneity assumption can be evaluated by testing for second-order serial correlation in the estimated residuals of the regression. Note that, since $\mathbb{E}(\Delta\epsilon_{prt}\Delta\epsilon_{prt-1}) = \mathbb{E}[(\epsilon_{prt}-\epsilon_{prt-1})(\epsilon_{prt-1}-\epsilon_{prt-2})] = -\,\mathbb{E}(\epsilon_{prt-1}^2) < 0$, negative first-order serial correlation is expected by construction.

The validity of the four-period lag used as an instrumental variable in this specific application to the German data is considered in Section 5.1.

3 Literature review

3.1 Methodologies

The approach for estimating the impact of immigration on different native skill groups detailed in Section 2 belongs to the "pure spatial" category described in Dustmann, Schönberg, and Stuhler (2016). The shared feature of studies that use this approach is their focus on the total immigration shocks to different regions, rather than the shocks to specific skill groups at the national level (the "national skill-cell approach") or in different regions (the "mixture approach"). The impact of this total immigration shock on different native skill groups is estimated using variation across regions and time.

Under the assumption of perfectly inelastic labor supplies for different skill groups, Dustmann, Schönberg, and Stuhler (2016) show that the pure spatial approach identifies the absolute impact of immigration on each native skill group, accounting for both the negative effect of increased competition from immigrants and the positive effects from complementarity between skill groups and the induced capital flows. The relative impact of immigration can then be identified by comparing the absolute impact for different skill groups. In contrast, the the national skill-cell and mixture approaches identify only the relative impact of immigration between skill groups but not the absolute impact, producing estimates with more limited interpretations.

A futher disadvantage of the national skill-cell and mixture approaches is that they produce estimates biased by immigrant downgrading. Since they focus on the immigration shocks to specific skill groups, these approaches assume that immigrants and natives with similar qualifications compete with each other. In reality, immigrants may instead compete with natives of lesser qualifications due to barriers to integration that force them to downgrade into lower-skilled occupations. The pure spatial approach circumvents this issue by considering the impact of only the total immigration shock. Due to this and the other advantages of the pure spatial approach, and to avoid comparing estimates with different interpretations, I review the results of only studies that use this approach in Section 3.2, most of whose estimates are summarized in Table 1 of Dustmann, Schönberg, and Stuhler (2016).

It is worth highlighting the differences between the particular pure spatial approach in Dustmann, Frattini, and Preston (2013) that I follow and the one classified in Dustmann, Schönberg, and Stuhler (2016). While the former uses a first-difference estimator that eliminates regional fixed effects but controls for time fixed effects by estimating them, the latter describes a difference-in-difference estimator that eliminates both the regional and time fixed effects. If the time fixed effects estimated by the first-difference estimator are biased due to a small sample size, the difference-in-difference estimator may produce less biased estimates of the impact of immigration. The difference-in-difference estimator also eliminates skill group-specific time trends, an additional control which the first-difference estimator ignores. Differential time trends across native wage percentiles in the estimated residuals from first-difference regressions would therefore be concerning.

The most unique feature of the approach in Dustmann, Frattini, and Preston (2013) is their definition of skill groups by percentiles of the native wage distribution. This definition does not allow for elastic labor supplies, let alone heterogenous elasticities across skill groups: native employment levels for different skill groups cannot be measured according to this definition, and the differential impact of immigration on native employment cannot be estimated. In contrast, studies like Dustmann, Schönberg, and Stuhler (2017) allow for elastic labor supplies by estimating the joint responses of native wages and employment to immigration for different skill groups defined by education and experience categories. This distinction between studies that define native skill groups by percentiles and those that do so by workers' observable characteristics is important to consider, since estimates from the latter of the impact of immigration on native wages may be moderated by the possibility of native employment responses.

3.2 Estimates

It is natural to first review the results in Dustmann, Frattini, and Preston (2013), since theirs will be the most directly comparable to my own. The authors consider the impact of an immigration wave that increased the immigrant-to-native ratio in the United Kingdom's labor force by about 3 percentage points between 1997 and 2005, defining immigrants as foreign-born workers who arrived to the country within 2 years from the time of measurement. They find that a 1 percentage-point increase in the immigrant-to-native ratio decreases the 5th and 10th percentiles of native wages by 0.67% and 0.52%, respectively, while increasing the 50th and 90th percentiles by 0.66% and 0.41%. The corresponding increases in the 50-5 and 50-10 native wage differentials are 1.3% and 1.1%, while the 90-50 differential barely changes. Immigration thus increased inequality in the bottom half of the native wage distribution without much impact on its upper half.

The authors also find a 0.4% increase in average wages, consistent with their theoretical prediction of a weakly positive impact on the average based on the same model derived in Section 2.1, given an elastic capital supply. Dustmann, Fabbri, and Preston (2005) use the same econometric approach and the same data for the United Kingdom, but for the years from 1983 to 2000, to estimate a 0.9% increase in average wages in response to a 1 percentage-point increase in the immigrant-to-native ratio. Overall, immigration during the late twentieth and early twenty-first century helped the median and average worker in the United Kingdom, but hurt the poorest ones.

The research on the impact of immigration on native wages in the United States features similar themes. For the period from 1970 to 1980, Altonji and Card (1991) show that a 1 percentage-point increase in the immigrant share of the population decreases the average wages of low-educated native workers by 1.2%, and by even more so for low-educated black workers. For the period from 1990 to 2010, following Dustmann, Frattini, and Preston (2013) but defining skill groups using an index of occupational skills, Sharpe and Bollinger (2020) estimate a corresponding 0.55% decrease in wages at the 10th percentile of the native skill distribution and a 0.36% increase at the 80th percentile.

Concerning the impact of immigration on average native wages in the United States, whereas Boustan, Fishback, and Kantor (2010) find no statistically significant impact between 1935 and 1940, Card (2007) estimates that a 1 percentage-point increase in the immigrant share of the population increased the average by 0.4% between 1980 and 2000. Like in the United Kingdom at the turn of the millennium, low-skilled workers in the twentieth-century United States were hurt by immigration, higher-skilled workers benefitted from it, and its impact on the average native was neutral-to-positive.

A closely studied natural experiment in the United States context is the case of the Mariel Boatlift, by which about 125,000 Cuban immigrants arrived to the city of Miami, Florida in 1980. The pioneering study (Card 1990) finds minimal negative impact of this sudden immigration wave on the wages of non-Cubans based on difference-in-difference estimates. Recent studies (Borjas 2017; Peri and Yasenov 2019) use synthetic control methods to reevaluate this case, reaching starkly contrasting conclusions about the impact on the wages of non-Cuban high-school dropouts. While the former estimates a decrease of greater magnitude than any previously discussed, the latter finds no impact, a difference whose reconciliation is beyond the scope of this paper.

A similar natural experiment has been the focus of most existing studies of the impact of immigration on native wages in Germany: the fall of the Berlin Wall and the reunification of Germany. Dustmann, Schönberg, and Stuhler (2017) evaluate the impact of a policy introduced in 1991, shortly after the Wall fell, that permitted workers from the Czech Republic to work in regions of Germany close to the German-Czech border. They estimate that a 1 percentage-point increase in the Czech share of the labor force decreases average native wages by only 0.13%, and by only 0.20% for unskilled workers. The corresponding decrease in native employment is 0.93%, an employment response which likely moderates the estimated negative impact on wages. Considering a related case, Prantl and Spitz-Oener (2020) evaluate the impact of East German immigration to West Germany. They find a negative effect on West German wages for workers in a competitive segment of labor markets, though their estimate based on the national skill-cell approach is not directly comparable to any of those previously discussed.

There are not yet any published studies on the impact on native wages of the immigration waves into Germany that began in 2011. However, Brunow and Jost (2020) find that immigrants from the Central and Eastern European countries that joined the European Union in the 2000s are less educated and work in lower-skilled occupations than native Germans and immigrants from the earlier member states. Since these Central and Eastern European immigrants have been the driving force behind the immigration wave that started in 2011, the authors' finding suggests that primarily lower-skilled natives have faced increased labor-market competition due to this wave.

4 Data and descriptive evidence

4.1 The SIAB

The data that I use to estimate the impact of immigration on native wages in Germany is from the Employee History segment of the Sample of Integrated Labor Market Biographies (SIAB), a 2% sample of administrative records for workers whose employment is subject to social security contributions or who are marginally employed. Documented in Antoni et al. (2019), this data is based on employer filings that carry severe legal penalties for misreporting and thereby ensure minimal errors. Given its accuracy and large sample size, the SIAB has been the gold standard for empirical studies of wage inequality in Germany since at least Dustmann, Ludsteck, and Schönberg (2009).

The data contains a variable for country of nationality, which is used to define workers with German nationality as natives and all others as immigrants. There is no variable for duration of residence in Germany by which recent immigrants can be distinguished from earlier ones, so the former's impact on native wages cannot be disentangled from that of the latter, a problem discussed in Section 2.3. However, since there is no reason to think that earlier immigrants moved in large numbers within Germany in recent decades, the changes in the immigrant-to-native ratio used to estimate equation (14) are likely driven by new arrivals. As long as the demographic trends among recent and earlier immigrants resemble each other, as discussed in Section 4.2, the estimated impact of immigration reflects mostly the impact of new arrivals.

All the years available for the new millennium—from 2000 to 2017—are initially considered. Though the immigration waves that motivate this study started only in 2011, the regional migration patterns from the prior decade may contain meaningful variation. Years prior to 2000 are ignored only to avoid complications arising from a major change to the data in 1999, when marginal part-time employees were first included. Though the data for the 2015-2017 period suffers from underreporting due to delayed filings by some employers, enough employers have already filed such that the samples currently available for these years are still representative of the population.

The SIAB reports workers' nominal daily wages but not their hours worked, making it impossible to compare wages between part-time and full-time workers. I therefore consider only workers labeled as full-time. A change to the data in 2011 led to a sharp increase in part-time observations compared to previous years before they restabilized in 2012. To use data from before 2012, I follow the procedure in Fitzenberger and Seidlitz (2020) to downweight full-time observations that are estimated to have a non-zero probability of being part-time. Appendix B.1 details the implementation of this procedure, while also elaborating on my other data-cleaning and sample-selection steps.

Using the authors' published Stata code as a starting point for my own, I restrict my sample to full-time workers who are between 25- and 55-years-old, work in the former states of West Germany (excluding Berlin), earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. The reason for distinguishing between West and East Germany is that their wage structures remain sufficiently different, even multiple decades since reunification, to make pooling them inappropriate when studying wage inequality.

I analyze the natural logarithm of real wages deflated using the Consumer Price Index. Throughout my analysis, observations are weighted according not only to their estimated full-time probability, but also to the fraction of the given year for which they are valid. If a worker has multiple employers during a given year, their annual wage is aggregated as a weighted average of their wages across employers, and their maximum education and occupational skill levels are retained retained. Observations with missing education are dropped.

The SIAB is right-censored at the threshold for social security contributions: in the average year, wages are censored for men and women at the 86th and 95th percentiles, respectively. Following the procedure in Gartner (2005), I estimate censored regressions to impute wages for these censored observations. Imputed wages are preferred to actual wages only for more sensible descriptive analysis, for example, the average wages by occupational skill levels reported in Table 2. To avoid misrepresentation, results along the native wage distribution are not reported above the 85th percentile. I forgo an analysis of the impact of immigration on average native wages, since censoring makes inaccurate the sample averages based on even imputed wages.

Finally, percentiles of the native wage distribution are calculated for each the 10 former states of West Germany in each of the 18 years considered, using regional samples that vary in size from 150 to 6,500 weighted immigrants and 2,500 to 65,000 weighted natives. Using these same weights, observations are aggregated to calculate immigrant-to-native ratios and demographic controls.

4.2 Age, gender, and education demographics

Per Figure 1, the start of the recent immigration waves in 2011 is visible in the data. After holding stable between 7.5% and 8.2% for the first decade of the new millennium, the immigrant share of the labor force increased by about 3 percentage points to over 10.5% between 2010 and 2017. Consistent with popular narratives, Figure 2 shows that this increase was driven primarily by an increase in immigration from Europe, and that there was an acceleration in immigration from the Middle East, Asia, and Africa in 2015.

Figure 1: The immigrant share of the labor force

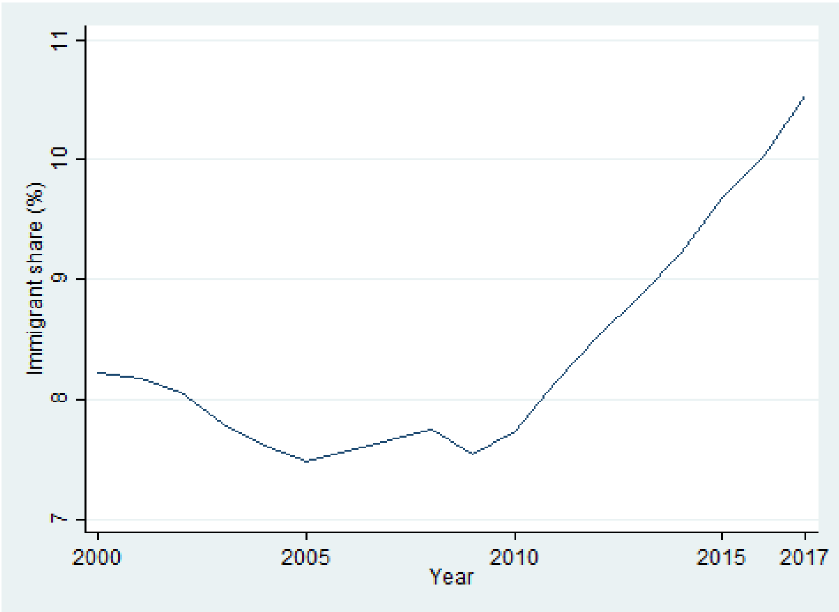
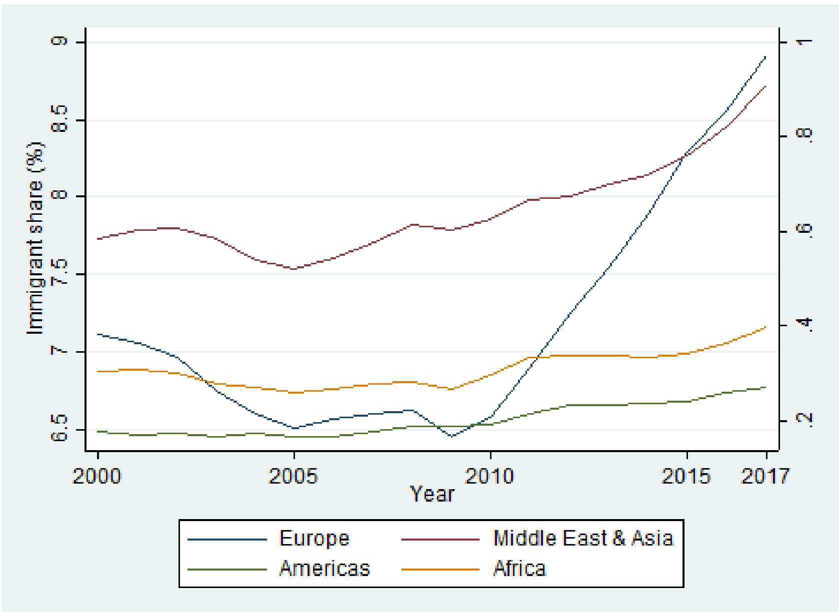


Figure 2: The immigrant share by region of nationality



Notes: Figure 1 plots the number of weighted immigrant workers as a fraction of the total weighted full-time labor force in the former states of West Germany (excluding Berlin) since 2000 based on the SIAB. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Figure 2 plots the same fraction but for immigrants with nationalities from countries in different regions of the world, measured on the left axis for Europe and on the right axis for all other regions. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).

Table 1: Age, gender, and education demographics

	Native			Immigrant		
	2000	2010	2017	2000	2010	2017
Average age	38.96	40.91	40.86	38.08	39.34	39.76
% Female	30.02	30.64	31.22	23.71	24.61	26.41
Education						
% Low	6.29	4.80	3.58	35.84	25.64	18.06
% Medium	79.27	76.55	72.88	57.05	59.94	59.94
% High	14.43	18.65	23.54	7.11	14.43	22.00

Notes: Displayed are the average age and female, low-, medium-, and high-educated shares of the weighted full-time native and immigrant labor forces in the former states of West Germany (excluding Berlin) in various years based on the SIAB. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).

As highlighted in Table 1, these immigration waves have not meaningfully changed immigrant age and gender demographics, nor have their demographics differed much from those of natives since 2000. The average age of native and immigrant workers has been around 40 throughout the 2000s, with only a small increase of about 2 years in both populations that occurred mostly between 2000 and 2010. The gap between natives and immigrants in terms of the women's share of the labor force is larger but still small, about 5-7 percentage points higher among natives than immigrants. The gap did, however, decrease faster between 2010 and 2017 than in the decade prior.

The differences between natives and immigrants in terms of education demographics are more stark. Since the reported education variable is missing for a large number of observations, I define education categories (low for completion of secondary school, medium for upper secondary school or vocational training, and high for university) based on the variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018), a variable which as of the latest version of the SIAB is now directly included in the data.

Table 1 demonstrates that, throughout the 2000s, immigrants have been more likely to be low-educated compared to natives and less likely to be either medium- or high-educated. While less than 7% of natives are low-educated in any year, almost 36% of immigrants were in 2000. Relatedly, the medium-educated share of the native labor force has been in the 70-80% range, whereas the equivalent share of the immigrant labor force has been in the 55-60% range.

However, as education levels have increased in both populations over time but faster among immigrants, these differences have narrowed. The already small low-educated share of natives has fallen even further alongside the medium-educated share, leading to an almost 10% increase in the high-educated share. In contrast, the medium-educated share of immigrants has remained relatively stable, while the high-educated share has grown substantially as the low-educated share has shrunk accordingly. The high-educated share of immigrants almost even matched that of natives in 2017.

This noticeable increase in education levels among immigrants is also not primarily explained by the recent immigration waves, since the increases between 2010 and 2017 occurred at similar rates to those between 2000 and 2010. Overall, Table 1 suggests that the trends in the age, gender, and education demographics of the new immigrants who began arriving starting in 2011 resemble the trends in those of the immigrant population of the prior decade.

4.3 Skills and downgrading

The lower education levels among immigrants compared to natives imply that immigrants are less skilled on average and more likely to work in lower-skilled occupations than natives. Table 2 confirms this reality: in 2017, the year in which their education levels were at their highest, immigrants were still almost 3 times as likely to work in unskilled or semi-skilled occupations as natives, and only 2/3 as likely to work in complex or highly complex ones. Analysis of the changes in this occupational skill distribution since 2000 is impossible, since they are based on a classification system introduced only at the end of 2011; the variable is available for prior years, but based on an imperfect conversion from an older system that thereby leads to substantial inaccuracies (Paulus and Matthes 2013). Evidence that this measure of occupational skill is a reasonable proxy for worker skill in 2017 is provided in the form of the average wage—across natives and immigrants—increasing with this measure.

These skill disparities between natives and immigrants are likely overstated due to downgrading by the latter: immigrants face disadvantages in the job-search process compared to natives (e.g. language barriers, discounted foreign qualifications, less developed social and professional networks) that may

Table 2: Occupational skill distributions

	Native	Immigrant	Average wage
Unskilled/semi-skilled	8.60	25.39	84.48
Skilled	56.44	51.44	109.64
Complex	17.87	9.64	145.89
Highly complex	17.09	13.53	167.71

Notes: Displayed are the occupational skill distributions of the weighted full-time native and immigrant labor forces in the former states of West Germany (excluding Berlin) in 2017 based on the SIAB, as well as the weighted-average real wages for each occupational skill level deflated into 2017 terms using the Consumer Price Index. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Occupational skill levels are defined according to the German Classification of Occupations 2010 described in Paulus and Matthes (2013). Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).

force them to work in occupations below their true skill level on first arrival to Germany, until they integrate into German society over time. Downgrading is likely to be more of an issue for recent immigrants than for ones who have been in the country for longer and had the time to better integrate. Given the nationality-based definition of immigrants used in this paper, the evidence below likely understates the extent of downgrading that new immigrants endure on arrival in Germany, while illustrating the integration barriers faced by the immigrant of the average duration of residence.

Table 3 provides evidence of immigrant downgrading in 2017 by showing that, at each education level, immigrants are overrepresented in lower-skill occupations compared to natives. Low-educated immigrants are almost as unlikely as low-educated natives to work in complex or highly complex occupations, but are more likely to do so in unskilled or semi-skilled ones. Similarly, medium-educated immigrants are almost as likely as natives to work in skilled occupations, but almost 3 times as likely to do so in unskilled or semi-skilled ones. Downgrading appears to be least prevalent among high-educated immigrants, since they are as likely—about 50%—as high-educated natives to work in highly complex occupations. 5% of them nevertheless still work in unskilled or semi-skilled occupations, whereas almost no high-educated natives do.

Following Dustmann, Frattini, and Preston (2013), I further compare the skill distributions of natives and immigrants and evaluate the extent of immigrant downgrading by ranking immigrants in the native wage distribution. In

Table 3: Occupational skill distributions by education

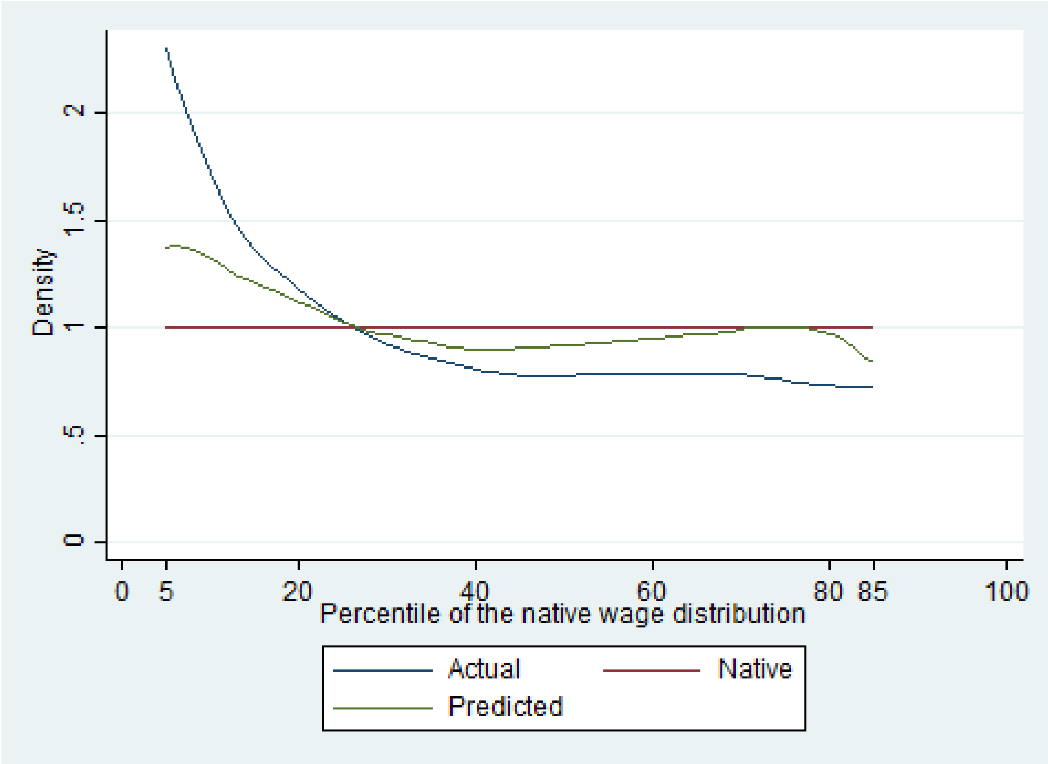
	Education					
	Low		Medium		High	
	Native	Immigrant	Native	Immigrant	Native	Immigrant
Unskilled/semi-skilled	35.76	48.52	9.65	25.78	1.20	5.37
Skilled	56.53	47.93	67.63	62.87	21.79	23.17
Complex	4.96	2.54	15.95	7.56	25.77	21.14
Highly complex	2.75	1.01	6.77	3.80	51.24	50.32

Notes: Displayed for different education categories are the occupational skill distributions of the weighted full-time native and immigrant labor forces in the former states of West Germany (excluding Berlin) in 2017 based on the SIAB. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university. Occupational skill levels are defined according to the German Classification of Occupations 2010 described in Paulus and Matthes (2013). Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).

the sense described in Section 2.2, workers' ranks in the native wage distribution are used to define different skill groups. The rank of a given immigrant in the native wage distribution in a given year is calculated as the percentage of (weighted) natives who earn a lower wage during that year. After computing these ranks, a kernel estimate of the density of their log odds—to account for the boundedness of rank from 0 to 1—can be fitted, appropriately retransformed, and compared to the density of natives' ranks in their own distribution, which always equals 1 by definition. Appendix B.2 details the implementation of this procedure.

Figure 3 plots the resulting density of immigrants relative to natives for the full-time labor force in the former states of West Germany during the 2011-2017 period covering the recent immigration waves. It shows that immigrants are overrepresented in the native wage distribution below about the 25th percentile and underrepresented above it. These results are consistent with those in Table 2: just as immigrants are overrepresented in lower-skilled occupations compared to natives, they are also overrepresented in lower-wage occupations. The blue line traces the estimated relative densities at different percentiles of the native wage distribution (i.e. $\widehat{\zeta}_p$) discussed in Section 2.3.

Figure 3: The density of immigrants relative to natives



Notes: Plotted are weighted kernel estimates of immigrants' ranks in the native wage distribution (i.e. the percentage of weighted natives who earn a lower wage) in the former states of West Germany (excluding Berlin) for each year from 2011 to 2017 based on the SIAB. Included are full-time workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Predicted wages for immigrants are based on weighted gender- and year-specific regressions of the natural logarithm of real native wages on a constant, age categories (25-30, 31-35, 36-40, 41-45, 46-50, 51-55), education categories (low, medium, high), and their interactions, clustering standard errors by age-education groups. To the predictions based on these regressions are then added error terms drawn from a normal distribution with a mean of zero and variance equal to the estimated forecast variance. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Nominal wages are deflated into 2017 terms using the Consumer Price Index. Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university.

To explore whether the overrepresentation of immigrants in lower-wage occupations is partially explained by downgrading, I predict the wages that immigrants would earn if they enjoyed the same returns to their observable characteristics as natives. Separately for each gender in each year, I estimate a (weighted) regression of only *natives'* wages on a constant, 6 age categories (25-30, 31-35, and every distinct 5-year span up to 55), 3 education categories, and their interactions, clustering standard errors by age-education groups. I then predict *immigrants'* wages using their age, gender, and education characteristics, adding a normally distributed error term with a mean of zero and variance equal to the estimated forecast variances from these regressions. I finally compute immigrants' predicted ranks in the native wage distribution and fit a kernel estimate of them as I did for their actual ranks.

The green line in Figure 3 traces the resulting predicted relatives density. It shows that, if immigrants were rewarded for their observable characteristics just like natives, they would be much less overrepresented below the 25th percentile of the native wage distribution, and barely underrepresented above it. These results are consistent with those in Table 3: since immigrants are overrepresented in lower-skilled occupations compared to natives of similar qualifications, they compete more with lower-wage natives than they should. Immigrant downgrading thus amplifies the increased labor-market competition faced by lower-skilled natives due to immigration.

It is worth noting the resemblance of Figure 3 to Panel C of Figure 1 in Dustmann, Schönberg, and Stuhler (2016), which plots the same values but for all of Germany, the year 2000, and only immigrants who arrived within 2 years from the time of measurement. The shape of the density of immigrants relative to that of natives did not change much between 2000 and the 2011-2017 period, though the disparities became less amplified. For example, the relative density of immigrants at the 5th percentile of native wages was 4 in 2000, compared to only about 2.5 during the 2011-2017 period. The extent of immigrant downgrading also remained similar: in both cases, the predicted relative density explains roughly half of the amplitude of the actual relative density's deviations from the native density of 1.

As a final contribution to the descriptive evidence of immigrant downgrading, I show that natives earn a wage premium over immigrants of identical observable characteristics, the majority of which is explained by immigrants' selection into lower-wage occupations and industries. I first estimate a (weighted) regression of all full-time workers' wages on a constant, age categories, education categories, their interactions, year fixed effects, and a dummy variable indicating German nationality. The estimated coefficient on this dummy variable is 0.098. I then estimate the same regression while also controlling for 10 occupation categories, 12 industry categories, and their interactions, resulting in an estimated coefficient of 0.047. Over 50% of the wage gap between natives and immigrants is thus explained by the employment of immigrants in lower-wage occupations and industries.

5 Main results

5.1 The impact of immigration on native wages

The primary objective of this paper is to estimate equation (14) to determine the impact of immigration along the distribution of native in the 10 former states of West Germany from 2000 to 2017. The main results of various implementations of this estimation are captured in Table 4.

Columns (1) and (2) of Table 4 report the estimated coefficient on the immigrant-to-native ratio from first-difference OLS regressions of every 5th percentile of the native wage distribution on this ratio and controls, using the full sample period. Column (1) controls only for year fixed effects, while column (2) also controls for demographics, specifically, the average ages of natives and immigrants and the natural logarithms of the ratios of high- and medium-educated natives to low-educated ones. The estimates in column (2) are of larger magnitude than those in column (1) for most percentiles, but only by a small amount that rarely exceeds 0.1, suggesting that changes in the immigrant-to-native ratio are not too correlated with changes in the demographics of natives and immigrants. The estimated standard errors are clustered by state. Though 10 clusters is usually too small for consistent estimation of standard errors, regional clustering is appropriate in this case like in Dustmann, Frattini, and Preston (2013), who consider only 17 regions.

These initial OLS estimates of the impact of immigration are statistically distinguishable from zero only at the middle and very end of the native wage distribution, but their signs along the entire distribution are nevertheless striking: they are negative below the 25th percentile and positive above it, mirroring the pattern in the relative density estimates in Figure 3. There is thus preliminary evidence in support of the theory from Section 2.1 that

Table 4: The impact of immigration along the native wage distribution

	OLS		IV		OLS starting in 2011		IV starting in 2011	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5th percentile	-0.175	-0.271	0.333	-0.541	-0.818	-1.040**	-0.434	-1.161**
	(0.439)	(0.406)	(1.829)	(1.353)	(0.489)	(0.385)	(0.724)	(0.485)
10th percentile	-0.180	-0.294	0.894	0.505	-0.517	-1.131*	-0.356	-0.823
	(0.462)	(0.393)	(1.171)	(1.042)	(0.708)	(0.551)	(0.470)	(0.455)
15th percentile	-0.074	–0.171	1.378	1.706	-0.186	-0.462	-0.020	-0.305
	(0.147)	(0.150)	(1.413)	(1.910)	(0.333)	(0.265)	(0.539)	(0.543)
20th percentile	-0.139	-0.144	2.137	2.015	-0.322	-0.434	0.124	-0.201
	(0.197)	(0.145)	(1.483)	(1.632)	(0.368)	(0.304)	(0.512)	(0.445)
25th percentile	0.022	0.143	2.455	2.754	-0.008	0.151	0.571	0.450
	(0.195)	(0.174)	(1.571)	(2.034)	(0.254)	(0.235)	(0.464)	(0.408)
30th percentile	0.186	0.245*	2.504	3.021	0.347*	0.452*	0.829*	0.759
	(0.145)	(0.115)	(1.452)	(2.125)	(0.177)	(0.201)	(0.451)	(0.480)
35th percentile	0.421***	0.448***	2.694*	3.096	0.564***	0.848***	1.098***	1.089***
	(0.063)	(0.085)	(1.335)	(1.849)	(0.121)	(0.164)	(0.334)	(0.330)
40th percentile	0.304**	0.321**	2.574*	2.698	0.407**	0.638***	0.958**	0.784**
	(0.109)	(0.108)	(1.293)	(1.585)	(0.160)	(0.170)	(0.335)	(0.262)
45th percentile	0.359***	0.398***	2.735*	2.883	0.485***	0.671***	1.059**	0.877***
	(0.083)	(0.095)	(1.411)	(1.686)	(0.146)	(0.183)	(0.457)	(0.253)
50th percentile	0.415**	0.457**	2.614	3.137	0.571**	0.737**	1.103*	1.103**
	(0.131)	(0.150)	(1.469)	(2.098)	(0.191)	(0.269)	(0.510)	(0.463)
55th percentile	0.320**	0.426***	2.329	2.962	0.379*	0.691***	1.158*	1.156*
	(0.106)	(0.086)	(1.316)	(1.974)	(0.168)	(0.200)	(0.580)	(0.513)
60th percentile	0.344**	0.453***	2.318	2.023	0.406	0.676***	1.241**	0.877*
	(0.145)	0.116	(1.360)	(1.502)	(0.260)	(0.160)	(0.546)	(0.405)
65th percentile	0.331	0.426**	1.989	1.299	0.443	0.751**	1.030**	0.580
	(0.189)	(0.143)	(1.185)	(1.272)	(0.385)	(0.234)	(0.397)	(0.438)
70th percentile	0.316	0.416**	1.745	0.997	0.323	0.562***	0.811*	0.416
	(0.204)	(0.169)	(1.188)	(1.340)	(0.275)	(0.167)	(0.442)	(0.617)
75th percentile	0.331	0.403	1.645	1.786	0.270	0.401	0.801	0.684
	(0.322)	(0.300)	(1.226)	(1.673)	(0.249)	(0.266)	(0.530)	(0.724)
80th percentile	0.304	0.219	1.438	1.670	0.273	0.473	0.608	0.577
	(0.216)	(0.223)	(0.914)	(1.383)	(0.225)	(0.325)	(0.591)	(0.777)
85th percentile	0.336**	0.288*	2.798*	3.184	0.547**	0.578*	1.237	1.389*
	(0.138)	(0.145)	(1.313)	(1.752)	(0.232)	(0.280)	(0.774)	(0.742)
F-statistic			3.545	3.555			12.49	54.72
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	170	170	140	140	70	70	70	70

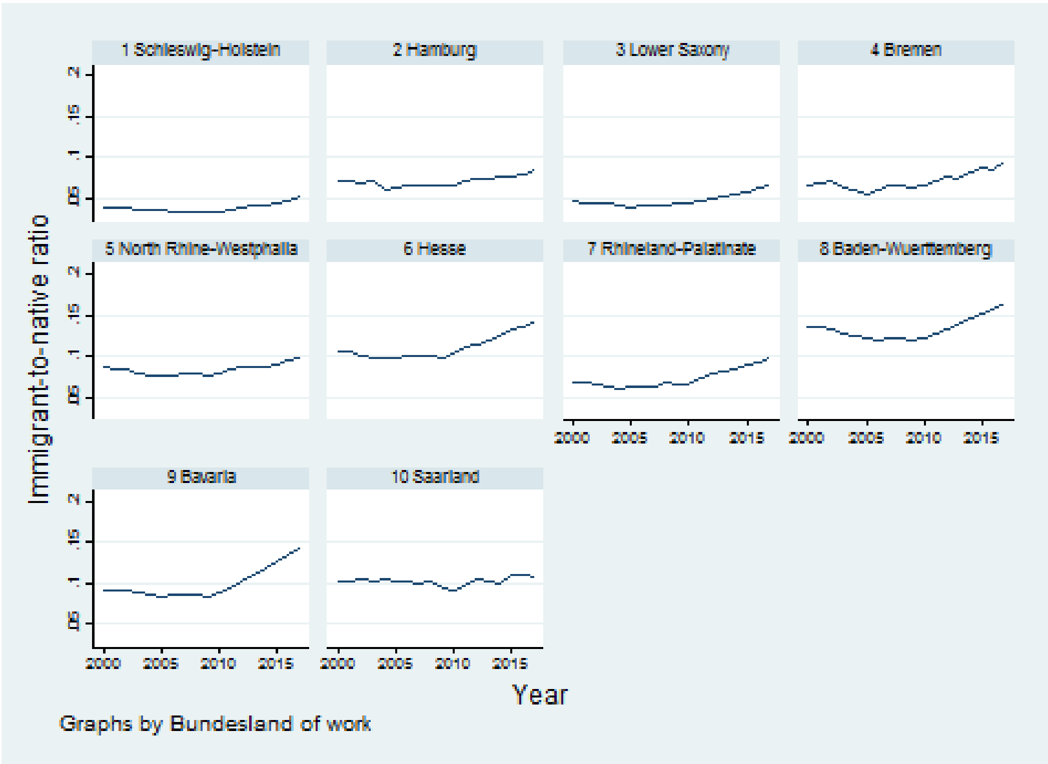
Notes: Displayed are the estimated coefficient on the immigrant-to-native ratio in first-difference regressions of different percentiles of native wages on this ratio. The estimated standard errors, clustered by state, are in parentheses. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Columns (1)-(4) are for the 2001-2017 period, while columns (5)-(8) are for the 2011-2017 period. The instrumental variable for the annual change in the ratio used in the IV regressions is its four-year lagged level. The *F*-statistics reported for the IV regressions is for the significance of excluded instruments in the first-stage regressions. Demographic controls include the average ages of natives and immigrants and the natural logarithms of the ratios of high- and medium-educated natives to low-educated ones. All variables are constructed using the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018).

immigration decreases natives wages for skill groups in which immigrants are overrepresented compared to natives and increases them for groups in which immigrants are underrepresented, assuming a perfectly elastic capital supply and the rank insensitivity of native wages to immigration. The first assumption is reasonable in this case, since capital can likely move sufficiently freely between the different states of Germany to approximate a perfectly elastic capital supply, given net inflows of immigrants that are not so massive that they push the national economy to its productive limits. The second assumption, captured in condition (7), is verified empirically in Section 5.2.

As discussed in Section 2.4, the OLS estimates are likely to be biased due to measurement error and immigrant selection into states with higher wage growth, biases which can be addressed by using a multi-period lagged level of the immigrant-to-native ratio as an instrumental variable for the change in this ratio. Choosing a lag that should be large enough to address this second bias but small enough to not shrink the sample period too drastically, I use a four-year lagged instrument that reduces the number of observations available for estimation by almost a fifth. The results from IV regressions without and with the aforementioned demographic controls are respectively reported in columns (3) and (4) of Table 4. The F-statistic for the significance of excluded instruments in the first-stage regression is about 3.5 in both cases, below the rule-of-thumb threshold of 10 above which the instrument could comfortably be deemed relevant. Due to this irrelevance, these initial IV estimates of the impact of immigration are nonsensically large and should be ignored.

A reason for the instrument's irrelevance using the full sample period may be that, prior to the start of the immigrations waves in 2011, movement by recent immigrants explains almost none of the variation in changes in the immigrant-to-native ratio. Instead, this variation may reflect movement by earlier immigrants, or it may even be too small to be anything but noise: as shown in Figure 4, the ratio barely changed before 2011 in any of the former West German states. In either case, the theoretical justification for the relevance of the instrument—that recent immigrants move to regions with higher concentrations of existing immigrants—breaks down, since the variation is not driven by the movement of recent immigrants. In pursuit of a relevant instrument that can be used to address the biases in the OLS estimates, it may be necessary to restrict the sample to the period starting in 2011, after which this variation is likely to be driven by recent immigrants.

Figure 4: The immigrant-to-native ratio by state of work



Note: Plotted are the immigrant-to-native ratios for each of the 10 former states of West Germany (excluding Berlin) between 2000 and 2017. The ratio is constructed using the SIAB, based on weighted full-time workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and their probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).

Columns (5) and (6) of Table 4 report the OLS estimates of the impact of immigration on native wages without and with demographic controls, respectively, for this restricted sample period from 2011 to 2017. These estimates are mostly of noticeably larger magnitude than their full-sample equivalents, suggesting a stronger correlation between changes in the immigrant-to-native ratio and wage growth after the start of the recent immigration waves. This stronger correlation supports the hypothesis that the variation in these changes before 2011 were too small to be anything but noise and should therefore be ignored. The estimates controlling for demographic changes are again of larger magnitude than those ignoring these controls, but the differences between the two sets of estimates are greater than in the full-sample case. These greater differences suggest that changes in the immigrant-to-native ratio are more correlated with changing demographics after 2011 than before it. Though it comes at the cost of more than halving the sample size to only 70 observations, restricting the sample period to only the years covering the recent immigration waves appears to produce more meaningful estimates.

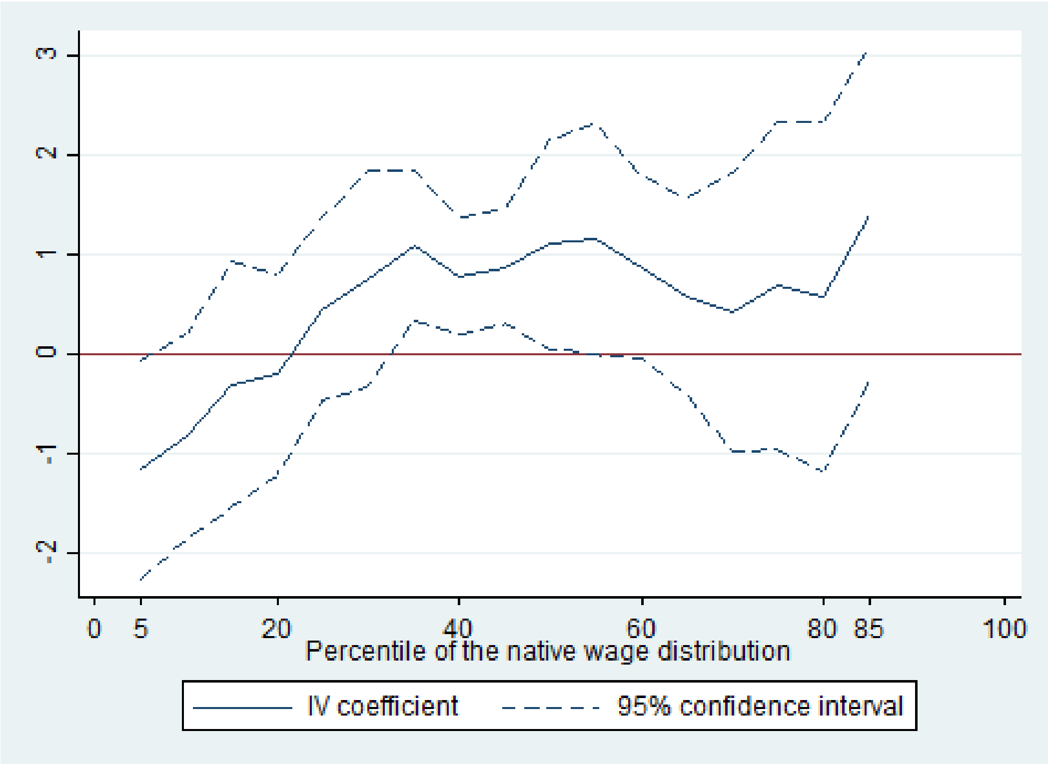
Finally, addressing the biases in these OLS estimates, columns (7) and (8) of Table 4 report the estimates from IV regressions for the restricted sample period from 2011 to 2017, using the four-year lagged immigrant-to-native ratio as an instrumental variable for the change in the ratio. The F-statistics for the significance of excluded instruments in the first-stage regressions are 12.5 when ignoring demographic changes and 54.7 when controlling for them. Since the former statistic remains too close for comfort to the rule-of-thumb threshold of 10, I focus my remaining analysis on the latter controlled case in which the instrument is clearly relevant. For evidence that it is also exogenous, consider the results in Appendix C.1 of Arellano-Bond tests for serial correlation in the estimated residuals of these IV regressions. For every percentile considered, the test fails to reject the null hypothesis of no second-order serial correlation at the 5% significance level. The test also fails to reject the absence of first-order serial correlation in the estimated residuals, a surprising result since negative correlation is expected by construction. The failure to reject first-order serial correlation may be explained by the small sample size.

A comparison of the estimates between columns (6) and (8) of Table 4 shows that the IV estimates of the impact of immigration along the distribution of native wages are of larger magnitude than the OLS estimates at most of the percentiles considered. Since the instrument addresses both the downward attenuation bias due to measurement error and the upward bias due to immigrant selection into states with higher wage growth, the mostly larger magnitudes of the IV estimates suggest that measurement error is the greater source of bias. The OLS estimates are also statistically significant at the 5% level at more percentiles than the IV estimates. The latter's apparently lower efficiency may be due to the small sample size of 70 observations, since IV estimation can suffer from its own issues of biasedness and inefficiency in such a finite-sample setting. I proceed with my analysis abstracting from the issues.

Figure 5 plots the IV estimates of the impact of immigration along the native wage distribution, as well as the 95% confidence intervals. It reinforces the preliminary observation made based on the OLS estimates using the full sample period: the impact remarkably mirrors the relative density of immigrants in Figure 3, negative below the first quartile percentile where immigrants are overrepresented and positive above it where they are underrepresented. Focusing on the percentiles where the estimates are statistically significant, a 1 percentage-point increase in the immigrant-to-native ratio leads to a 1.2% decrease in the 5th percentile of natives wages, and increases ranging from 0.8% to 1.1% between the 35th through 50th percentiles. Given these estimates and the about 3 percentage-point increase in the immigrant-to-native ratio after 2010, as reported in the summary statistics in Appendix C.2, the recent immigration waves increased the 50-5 wage differential in the average West German state by slightly more than 6%, a considerable increase in inequality in the bottom half of the native wage distribution.

It is worth highlighting the greater magnitudes of these estimates compared to most of those in the existing literature reviewed in Section 3.2. A reason for these greater magnitudes at the 5th and 50th percentiles compared to Dustmann, Frattini, and Preston (2013) may simply be that immigrants are more overrepresented in the bottom of the native wage distribution in Germany than they are in the United Kingdom. The relative density of immigrants in Germany peaks at almost 2.5 at the 5th percentile in Figure 3, at which it equals only 1.5 for the United Kingdom based on Figure 1 of the authors' paper. The much larger negative impact on lower-skilled native wages compared to Dustmann, Schönberg, and Stuhler (2017) may be due to the definition of skill groups by native wage percentiles, precluding the estimation of employment effects that could moderate the estimated impact on wages.

Figure 5: The impact of immigration along the native wage distribution



Notes: Plotted are the estimated coefficients on the immigrant-to-native ratio in first-difference IV regressions of different percentiles of native wages on this ratio, year fixed effects, and demographic controls for the 2011-2017 period. The estimated 95% confidence intervals are also plotted, based on standard errors clustered by state. The estimates match those in column (8) of Table 4. The instrumental variable for the annual change in the ratio used in the IV regressions is its four-year lagged level. Demographic controls include the average ages of natives and immigrants and the natural logarithms of the ratios of high- and medium-educated natives to low-educated ones. All variables are constructed using the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university.

5.2 The elasticity of substitution

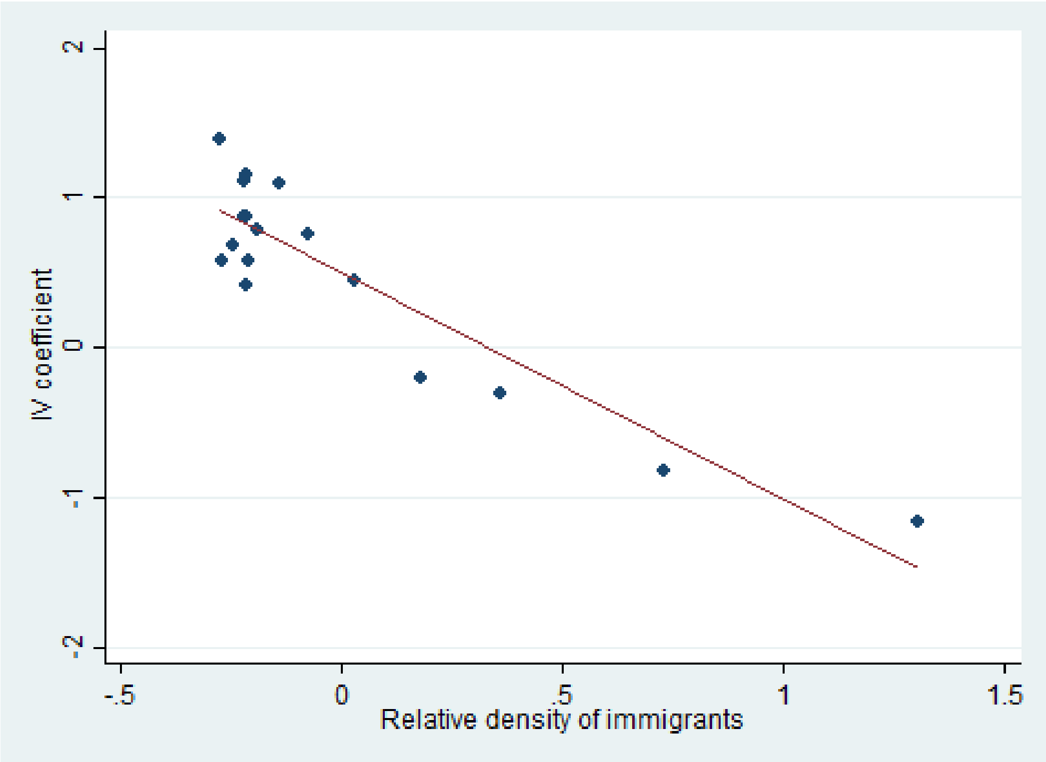
I now follow the procedure detailed in Section 2.3 to estimate the structural parameter σ and thereby the elasticity of substitution between different skill groups of workers. I fit an OLS regression of the IV estimates of the impact of immigration on native wages at every 5th percentile up to the 85th on the estimates of the relative density of immigrants at the same percentiles, as well as a constant. The former estimates $\widehat{\gamma}_p$ are displayed in column (8) of Table 4, while the latter $\widehat{\zeta_p}$ plus 1 are plotted in the blue line in Figure 3. As discussed in Appendix B.2, the plotted estimates are subtracted by 1 for clarity of interpretation—a value greater than 1 means immigrants are over-represented compared to natives at a given percentile, while a value less than 1 means they are underrepresented—with no impact on the fitted regression.

Figure 6 plots these estimates as well as this fitted regression line, whose slope coefficient is about -1.5. Given $\gamma_p = (\sigma - 1)\zeta_p$ from Section 2.2, the estimated structural parameter is $\widehat{\sigma} = -0.5$, and the estimated elasticity of substitution is $\frac{1}{1-\widehat{\sigma}} = \frac{2}{3}$. These structural estimates for Germany from 2011 to 2017 are similar to those in Dustmann, Frattini, and Preston (2013) for the United Kingdom from 1997 to 2005.

As discussed in Section 2.2, I can at last use these structural estimates to test a key assumption behind my results in Section 5.1: the rank insensitivity of native wages to immigration, a condition captured in condition (7) that is sufficient and necessary for a definition of different skill groups by different percentiles of the native wage distribution. Given the structural estimate $\widehat{\sigma} = -0.5$ and the 3 percentage-point increase in the immigrant-to-native ratio after 2010 such that $\widehat{m} = 0.03$, this condition can be empirically verified.

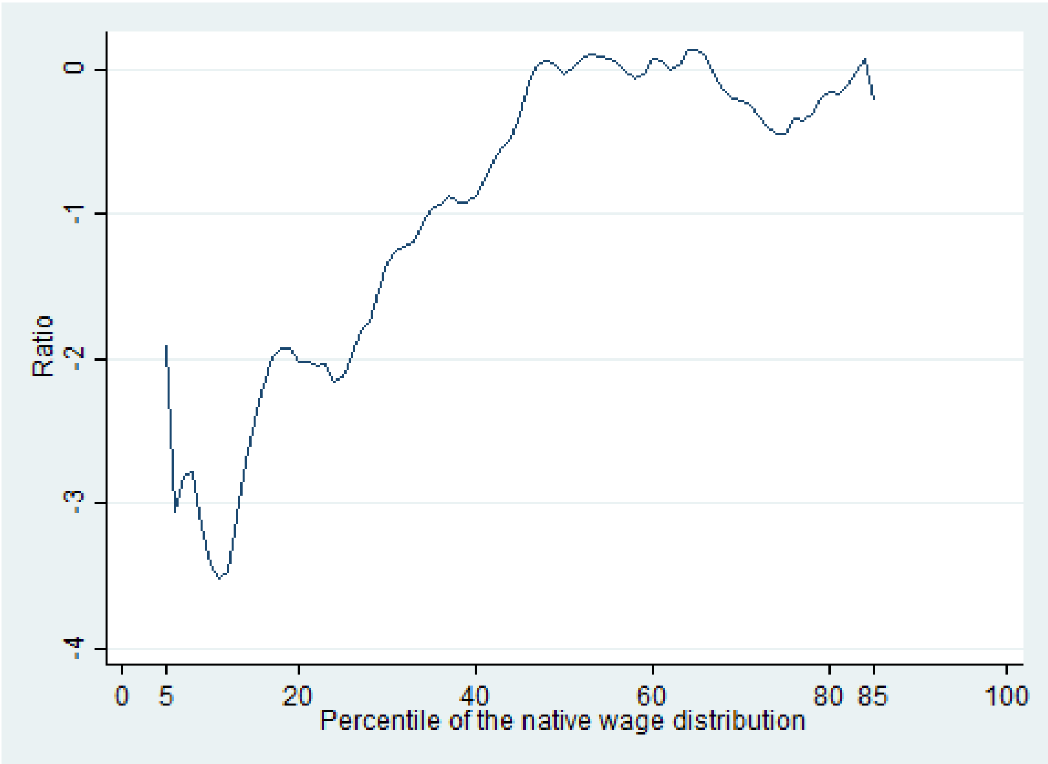
Based on these estimates, the ratio of the difference in the relative density of immigrants to the difference in log real wages between adjacent percentiles of the native wage distribution must equal greater than -22 across the entire distribution. Figure 7 plots this ratio for every pair of adjacent percentiles $p - 1$ and p from $p = 5$ to $p = 85$. It is never less than -4, so condition (7) is satisfied, and the definition of skill groups by percentiles validated.

Figure 6: Regression of the impact of immigration on native wages on the relative density of immigrants at different percentiles



Notes: Displayed on the vertical axis of this scatter plot are the estimated coefficients on the immigrant-to-native ratio in first-difference IV regressions of different percentiles of native wages on this ratio, year fixed effects, and demographic controls for the 2011-2017 period, from column (8) of Table 4. Displayed on the horizontal axis are weighted kernel estimates of immigrants' ranks in the native wage distribution (i.e. the percentage of weighted natives who earn a lower wage) for each year in the same period, from the blue line in Figure 3, but subtracted by 1 for clarity of interpretation: a value greater than 1 means immigrants are overrepresented compared to natives at a given percentile, while a value less than 1 means they are underrepresented. The red line traces the fitted regression of the former estimates on the latter. The instrumental variable for the annual change in the ratio used in the IV regressions is its four-year lagged level. Demographic controls include the average ages of natives and immigrants and the natural logarithms of the ratios of high- and medium-educated natives to low-educated ones. All data is from the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university.

Figure 7: Test of rank insensitivity to immigration



Notes: Displayed is the ratio of the difference in the estimated relative density of immigrants to the difference in log real wages between adjacent percentiles of the native wage distribution, which must equal greater than -22 across the entire distribution to validate the condition captured in condition (7) that immigration does not change native workers' ranks in their wage distribution. Relative density estimates are based on weighted kernel ones of immigrants' ranks in the native wage distribution (i.e. the percentage of weighted natives who earn a lower wage) for each year in the 2011-2017 period, from the blue line in Figure 3. All data is from the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Nominal wages are deflated into 2017 terms using the Consumer Price Index.

6 Conclusion

This paper has made three contributions. First, it is the first study of the impact on native wages of the immigration waves into Germany that started in 2011, finding that immigrants are overrepresented below the first quartile of the native wage distribution and underrepresented above it in the full-time West German labor force. Consistent with this pattern, it also finds a negative impact on the wages of the poorest natives and positive ones for those in the middle of the distribution. The recent immigration waves have therefore increased wage inequality in the bottom half of the native wage distribution in the average West German region between 2011 and 2017.

Second, based on these findings, this paper reinforces the main themes in the literature on the differential impact of immigration on native wages: the median native enjoys increases in their wages due to increased marginal productivity, whereas lower-skilled natives suffer wage decreases driven by increased labor-market competition. The intensity of the competition for lower-skilled occupations is partially explained by immigrant downgrading.

Lastly, this paper successfully applies the econometric approach in Dustmann, Frattini, and Preston (2013), whose definition of skill groups by percentiles of the native wage distribution allows for an analysis of the impact of immigration along the distribution, and consequently for precise claims about the impact on native wage inequality. The results from this application to Germany resemble those from the original one to the United Kingdom, providing empirical evidence for the theory motivating the approach.

This paper can be extended in two directions: robustness checks and further research. In the former case, variation across districts or commuting zones rather than states could be used. This narrower definition of regions would increase the sample size and allow for consistent estimation. It would also facilitate the inclusion of East Germany in this analysis or its reproduction on East German data. Another valuable extension would be to consider alternative instruments for the change in the immigrant-to-native ratio. Alternatives could be constructed using varying lags, official immigration statistics, or predicted immigrant inflows based on the procedure in Card (2001).

In terms of future research, this paper's findings inspire questions about the impact of the recent immigration waves on overall wage inequality in Germany. How would wage inequality have evolved without the 3 percentage-point increase in the immigrant share of the labor force caused by these waves?

What would the wage distribution have looked like without downgrading, namely, if immigrants' occupational skill distributions had matched those of similarly qualified natives? The literature on wage inequality offers a variety of methods through which to explore possible answers to these questions (Fortin, Lemieux, and Firpo 2011; Juhn, Murphy, and Pierce 1993; Lemieux 2006).

This paper offers a related insight: without the recent immigration waves, the differential between the 50th and 5th percentiles of the native wage distribution would have been 6% smaller in the average West German state. At first glance, this statement would appear to be a boon to the opponents of open borders and asylum, who might interpret it as evidence for their claim that lower-skilled immigrants and refugees are stealing jobs from working-class Germans (Gedmin 2019). This interpretation, however, ignores the reality of immigrant downgrading: the negative impact of immigration on the poorest natives is partially explained by German society's failure to integrate immigrants into the economy by facilitating their entry into occupations matching their qualifications. The native and immigrant working classes are collectively hurt by the barriers to integration faced by the latter in Germany, while their discounted labor benefits the native middle and capitalist classes. Blame not the immigrant, therefore, but the institutions responsible for those barriers.

Bibliography

Altonji, Joseph and David Card (1991). “The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives”. In: *Immigration, Trade, and the Labor Market*. National Bureau of Economic Research, pp. 201–234.

Antoni, Manfred et al. (2019). *Sample of Integrated Labour Market Biographies (SIAB) 1975-2017*. Tech. rep. The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Bertoli, Simone, Herbert Brücker, and Jesús Fernández-Huertas Moraga (Jan. 2013). *The European Crisis and Migration to Germany: Expectations and the Diversion of Migration Flows*. IZA Discussion Papers 7170. Institute of Labor Economics (IZA).

Borjas, George J. (Oct. 2017). “The Wage Impact of the *Marielitos*: A Reappraisal”. In: *ILR Review* 70.5, pp. 1077–1110. ISSN: 0019-7939, 2162-271X. DOI: 10.1177/0019793917692945.

Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor (Oct. 2010). “The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression”. In: *Journal of Labor Economics* 28.4, pp. 719–746. DOI: 10.1086/653488.

Brunow, Stephan and Oskar Jost (2020). *On the Foreign to Native Wage Differential in Germany: Does the Home Country Matter?* IAB Discussion Paper. Institute for Employment Research (IAB).

Card, David (Jan. 1990). “The Impact of the Mariel Boatlift on the Miami Labor Market”. In: *ILR Review* 43.2, pp. 245–257. ISSN: 0019-7939, 2162-271X. DOI: 10.1177/001979399004300205.

— (Jan. 2001). “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration”. In: *Journal of Labor Economics* 19.1, pp. 22–64. ISSN: 0734-306X, 1537-5307. DOI: 10.1086/209979.

Card, David (2007). *How Immigration Affects U.S. Cities*. CReAM Discussion Paper Series 0711. Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Dustmann, C., T. Frattini, and I. P. Preston (Jan. 2013). "The Effect of Immigration along the Distribution of Wages". In: *The Review of Economic Studies* 80.1, pp. 145–173. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rds019.

Dustmann, Christian, Francesca Fabbri, and Ian Preston (Nov. 2005). "The Impact of Immigration on the British Labour Market". In: *The Economic Journal* 115.507, F324–F341. ISSN: 0013-0133, 1468-0297. DOI: 10.1111/j.1468-0297.2005.01038.x.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (May 2009). "Revisiting the German Wage Structure". In: *Quarterly Journal of Economics* 124.2, pp. 843–881. ISSN: 0033-5533, 1531-4650. DOI: 10.1162/qjec.2009.124.2.843.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler (Nov. 2016). "The Impact of Immigration: Why Do Studies Reach Such Different Results?" In: *Journal of Economic Perspectives* 30.4, pp. 31–56. ISSN: 0895-3309. DOI: 10.1257/jep.30.4.31.

— (Feb. 2017). "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment". In: *The Quarterly Journal of Economics* 132.1, pp. 435–483. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjw032.

Fitzenberger, Bernd and Arnim Seidlitz (Dec. 2020). "The 2011 Break in the Part-Time Indicator and the Evolution of Wage Inequality in Germany". In: *Journal for Labour Market Research* 54.1, p. 1. ISSN: 2510-5019, 2510-5027. DOI: 10.1186/s12651-019-0265-0.

"Five Years after Arrival, Germany's Refugees Are Integrating" (Aug. 2020). In: *The Economist*. ISSN: 0013-0613.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011). "Decomposition Methods in Economics". In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1–102. ISBN: 978-0-444-53450-7. DOI: 10.1016/S0169-7218(11)00407-2.

Gartner, Hermann (June 2005). *The Imputation of Wages above the Contribution Limit with the German IAB Employment Sample*. Tech. rep. The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Gedmin, Jeffrey (July 2019). *Right-Wing Populism in Germany: Muslims and Minorities after the 2015 Refugee Crisis*. Tech. rep. Brookings.

Higgins, Matthew and Thomas Klitgaard (May 2019). *How Has Germany's Economy Been Affected by the Recent Surge in Immigration?* Tech. rep. The Liberty Street Economics Blog of the Federal Reserve Bank of New York.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993). "Wage Inequality and the Rise in Returns to Skill". In: *Journal of Political Economy* 101.3, pp. 410–442. ISSN: 00223808, 1537534X.

Lemieux, Thomas (2006). "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" In: *The American Economic Review* 96.3, pp. 461–498. ISSN: 00028282.

Paulus, Wiebke and Britta Matthes (Aug. 2013). *The German Classification of Occupations 2010 : Structure, Coding and Conversion Table*. Tech. rep. The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Peri, Giovanni and Vasil Yasenov (2019). "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift". In: *The Journal of human resources* 54.2, pp. 267–309. ISSN: 0022-166X.

Prantl, Susanne and Alexandra Spitz-Oener (Mar. 2020). "The Impact of Immigration on Competing Natives' Wages: Evidence from German Reunification". In: *The Review of Economics and Statistics* 102.1, pp. 79–97. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/rest_a_00853.

Sharpe, Jamie and Christopher R. Bollinger (Oct. 2020). "Who Competes with Whom? Using Occupation Characteristics to Estimate the Impact of Immigration on Native Wages". In: *Labour Economics* 66, p. 101902. ISSN: 09275371. DOI: 10.1016/j.labeco.2020.101902.

Thomsen, Ulrich, Johannes Ludsteck, and Alexandra Schmucker (2018). "Skilled or Unskilled - Improving the Information on Qualification for Employee Data in the IAB Employee Biography". In: DOI: 10.5164/IAB.FDZM.1809.EN.V1.

Appendices

A Theoretical derivations

This appendix details the derivations behind all the theoretical results presented in Section 2.

A.1 The wage equation

To derive the first-order conditions stated in Section 2.1, first differentiate equation (1) with respect to each type of labor and to capital, also differentiating equation (2) as an intermediate step:

$$\begin{aligned} w_i = \frac{\partial Y}{\partial l_i} &= \frac{1}{s}\left[\beta H^s + (1-\beta)K^s\right]^{\frac{1}{s}-1} s\beta H^{s-1}\frac{\partial H}{\partial l_i} \\ &= \left[\beta H^s + (1-\beta)K^s\right]^{\frac{1}{s}-1} \beta H^{s-1}\frac{1}{\sigma}\left(\sum_{i=1}^{n}\alpha_i l_i^{\sigma}\right)^{\frac{1}{\sigma}-1} \sigma\alpha_i l_i^{\sigma-1} \\ &= \beta\alpha_i l_i^{\sigma-1}\left(\sum_{i=1}^{n}\alpha_i l_i^{\sigma}\right)^{\frac{1}{\sigma}-1} H^{s-1}\left[\beta H^s + (1-\beta)K^s\right]^{\frac{1}{s}-1} \\ &= \beta\alpha_i(\pi_i^N + m\pi_i^I)^{\sigma-1}\left(\frac{H}{L_N}\right)^{1-\sigma}\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^s\right]^{\frac{1}{s}-1} \end{aligned} \tag{A.1}$$

$$\begin{aligned} \rho = \frac{\partial Y}{\partial K} &= \frac{1}{s}\left[\beta H^s + (1-\beta)K^s\right]^{\frac{1}{s}-1} s(1-\beta)K^{s-1} \\ &= (1-\beta)\left(\frac{K}{H}\right)^{s-1}\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^s\right]^{\frac{1}{s}-1} \end{aligned} \tag{A.2}$$

Then take the natural logarithms of equations (A.1) and (A.2) to derive equations (3) and (4).

A.2 The impact of immigration on native wages

To derive the precise impact of immigration on each type's log wage stated in Section 2.1, first differentiate equations (3) and (4) with respect to m:

$$\begin{aligned}
\frac{\partial \ln w_i}{\partial m} &= (\sigma-1)\frac{\pi_i^I}{pi_i^N + m\pi_i^I} + (1-\sigma)\frac{\partial \ln H}{\partial m} + \frac{1-s}{s}\left(\frac{(1-\beta)sK^{s-1}H^{-s}}{\beta+(1-\beta)\left(\frac{K}{H}\right)^s}\right)\frac{\partial K}{\partial m} \\
&\qquad - \frac{1-s}{s}\left(\frac{(1-\beta)sK^{s}H^{-s-1}}{\beta+(1-\beta)\left(\frac{K}{H}\right)^s}\right)\frac{\partial H}{\partial m} \\
&= (\sigma-1)\frac{\pi_i^I}{\pi_i^N + m\pi_i^I} + (1-\sigma)\frac{\partial \ln H}{\partial m} \\
&\qquad + (1-s)(1-\psi)\left(\frac{\partial \ln K}{\partial m} - \frac{\partial \ln H}{\partial m}\right) && \text{(A.3)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln \rho}{\partial m} &= (s-1)\left(\frac{\partial \ln K}{\partial m} - \frac{\partial \ln H}{\partial m}\right) + (1-s)(1-\psi)\left(\frac{\partial \ln K}{\partial m} - \frac{\partial \ln H}{\partial m}\right) \\
&= -(1-s)\psi\left(\frac{\partial \ln K}{\partial m} - \frac{\partial \ln H}{\partial m}\right) && \text{(A.4)}
\end{aligned}$$

where $\psi = \frac{\beta H^s}{\beta H^s + (1-\beta)K^s}$ is the human capital share of output. The impact of immigration on human capital is:

$$\begin{aligned}
\frac{\partial \ln H}{\partial m} &= \frac{\partial \ln\left(L^N\left[\sum_{i=1}^n \alpha_i(\pi_i^N + m\pi_i^I)^\sigma\right]^{\frac{1}{\sigma}}\right)}{\partial m} \\
&= \sum_{i=1}^n \frac{\alpha_i \pi_i^I(\pi_i^N + m\pi_i^I)^{\sigma-1}}{\sum_{i=1}^n \alpha_i(\pi_i^N + m\pi_i^I)^\sigma} \\
&= \sum_{i=1}^n \frac{\alpha_i(\pi_i^N + m\pi_i^I)^\sigma}{\sum_{i=1}^n \alpha_i(\pi_i^N + m\pi_i^I)^\sigma}\frac{\pi_i^I}{\pi_i^N + m\pi_i^I} \\
&= \sum_{i=1}^n \omega_i \frac{\pi_i^I}{\pi_i^N + m\pi_i^I} && \text{(A.5)}
\end{aligned}$$

where the weights $\omega_i = \frac{\alpha_i(\pi_i^N + m\pi_i^I)^\sigma}{\sum_{i=1}^n \alpha_i(\pi_i^N + m\pi_i^I)^\sigma}$ are different types of labor's shares of human capital that satisfy $\sum_{i=1}^n \omega_i = 1$. Then define the elasticity of the supply of capital $\theta = \frac{\partial \ln K}{\partial \ln \rho}$ and multiply both sides of this definition by $\frac{\partial \ln \rho}{\partial m}$:

$$\frac{\partial \ln K}{\partial m} = \theta \frac{\partial \ln \rho}{\partial m} \qquad \text{(A.6)}$$

Next combine equations (A.4) and (A.6) to derive the impact of immigration on capital:

$$\begin{aligned}\frac{\partial \ln K}{\partial m} &= -(1-s)\psi\theta\left(\frac{\partial \ln K}{\partial m} - \frac{\partial \ln H}{\partial m}\right)\\ &= \frac{(1-s)\psi\theta}{1+(1-s)\psi\theta}\frac{\partial \ln H}{\partial m} \end{aligned} \tag{A.7}$$

Finally combine equations (A.3), (A.5), and (A.7) to derive equation (5):

$$\frac{\partial \ln w_i}{\partial m} = (\sigma - 1)\left(\frac{\pi_i^I}{\pi_i^N + m\pi_i^I} - \phi\sum_{i=1}^{n}\omega_i\frac{\pi_i^I}{\pi_i^N + m\pi_i^I}\right)$$

where $\phi = 1 + \left[\frac{(1-s)(1-\psi)}{1+(1-s)\psi\theta}\right]\frac{1}{1-\sigma}$ depends on the substitutability of human capital and capital as determined by the parameter s and of different types of labor as determined by the parameter σ, on the human capital share of output ψ, and on the capital supply elasticity θ. Note that $\phi = 1$ if human capital and capital are perfectly substitutable (i.e. $s = 1$), if the capital share is zero (i.e. $\psi = 1$), or if the supply of capital is perfectly elastic (i.e. $\theta = \infty$). The first two cases are unrealistic, but the last case can be approximated if the capital supply is sufficiently elastic, for example, if capital is mobile across regions and the national economy is not at its productive limits.

A.3 Rank insensitivity to immigration

To derive the sufficient and necessary condition for rank insensitivity to immigration stated in Section 2.2, consider the ranks of each type of labor in the native wage distribution before immigration. For example, assume that type i earns a greater wage than type j, such that equation (3) implies:

$$\ln\left(\frac{w_i^0}{w_j^0}\right) = \ln\left(\frac{\alpha_i}{\alpha_j}\right) + (\sigma-1)\ln\left(\frac{\pi_i^N}{\pi_j^N}\right) > 0 \tag{A.8}$$

Their ranks are preserved after immigration if:

$$\ln\left(\frac{w_i}{w_j}\right) = \ln\left(\frac{\alpha_i}{\alpha_j}\right) + (\sigma-1)\ln\left(\frac{\pi_i^N + m\pi_i^I}{\pi_j^N + m\pi_j^I}\right) > 0 \tag{A.9}$$

Note that conditions (A.8) and (A.9) suggest that different types of labors' ranks are a function of their relative productivities α_i and α_j, as well as of their relative supplies. Substitute condition (A.8) into the right-hand side of condition (A.9):

$$\ln\left(\frac{w_i}{w_j}\right) = \ln\left(\frac{w_i^0}{w_j^0}\right) + (\sigma-1)\ln\left(\frac{1+m\frac{\pi_i^I}{\pi_i^N}}{1+m\frac{\pi_j^I}{\pi_j^N}}\right) > 0 \tag{A.10}$$

Then combine conditions (A.8) and (A.10) and note that $\ln(1+x) \approx x$ for small x :

$$\ln\left(\frac{w_i}{w_j}\right) \geq (\sigma-1)\ln\left(\frac{1+m\frac{\pi_i^I}{\pi_i^N}}{1+m\frac{\pi_j^I}{\pi_j^N}}\right) \approx (\sigma-1)m\left(\frac{\pi_i^I}{\pi_i^N} - \frac{\pi_j^I}{\pi_j^N}\right) \tag{A.11}$$

Finally divide both sides of condition (A.11) by $\ln\left(\frac{w_i}{w_j}\right)$ and $(\sigma-1)m$ to derive condition (7).

A.4 Taylor approximation of the wage equation

The first-order Taylor approximation of equation (3) around $m = 0$ is:

$$\begin{aligned}\ln w_i &\approx \ln w_i^0 + \frac{\partial \ln w_i^0}{\partial m} m \\ &= \ln(\beta\alpha_i) + (\sigma-1)\ln \pi_i^N + \frac{1-\sigma}{\sigma}\ln\left(\sum_{i=1}^n \alpha_i (\pi_i^N)^\sigma\right) \\ &\quad + \left(\frac{1}{s}-1\right)\ln\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^s\right]\Bigg|_{m=0} + \frac{\partial \ln w_i^0}{\partial m} m \end{aligned} \tag{A.12}$$

Substitute equation (4) into the right-hand-side of expression (A.12):

$$\begin{aligned}\ln w_i &\approx \ln(\beta\alpha_i) + (\sigma-1)\ln \pi_i^N + \frac{1-\sigma}{\sigma}\ln\left(\sum_{i=1}^n \alpha_i (\pi_i^N)^\sigma\right) \\ &\quad + \ln\left(\frac{\rho_0}{1-\beta}\right) - (s-1)\ln\left(\frac{K}{H}\right)\Bigg|_{m=0} + \frac{\partial \ln w_i^0}{\partial m} m \end{aligned} \tag{A.13}$$

where $\rho_0 = \rho|_{m=0}$ is the rate of return on capital before immigration. Then solve for the second-to-last term on the right-hand side of expression (A.13) starting from equation (A.2):

$$\begin{aligned}
\frac{\rho}{1-\beta} &= \left(\frac{K}{H}\right)^{s-1}\left[\beta + (1-\beta)\left(\frac{K}{H}\right)^{s}\right]^{\frac{1}{s}-1} \\
\Longrightarrow \left(\frac{1}{\beta}\right)^{\frac{1}{s}-1}\frac{\rho}{1-\beta} &= \left(\frac{K}{H}\right)^{s-1}\left[1 + \frac{1-\beta}{\beta}\left(\frac{K}{H}\right)^{s}\right]^{\frac{1}{s}-1} \\
\Longrightarrow \frac{1}{\beta}\left(\frac{\rho}{1-\beta}\right)^{\frac{s}{1-s}} &= \left(\frac{K}{H}\right)^{-s}\left[1 + \frac{1-\beta}{\beta}\left(\frac{K}{H}\right)^{s}\right] \\
\Longrightarrow \frac{K}{H} &= \left[\frac{1}{\beta}\left(\frac{\rho}{1-\beta}\right)^{\frac{s}{1-s}} - \frac{1-\beta}{\beta}\right]^{-\frac{1}{s}} \qquad \text{(A.14)}
\end{aligned}$$

Next substitute equations (A.14) and (6) into expression (A.13):

$$\begin{aligned}
\ln w_i \approx{}& \ln(\beta\alpha_i) + (\sigma-1)\ln \pi_i^N + \frac{1-\sigma}{\sigma}\ln\left(\sum_{i=1}^{n}\alpha_i(\pi_i^N)^{\sigma}\right) \\
&+ \ln\left(\frac{\rho_0}{1-\beta}\right) + \frac{s-1}{s}\ln\left[\frac{1}{\beta}\left(\frac{\rho_0}{1-\beta}\right)^{\frac{s}{1-s}} - \frac{1-\beta}{\beta}\right] \\
&+ (\sigma-1)\left(\frac{\pi_i^I}{\pi_i^N} - \phi\sum_{i=1}^{n}\omega_i^0\frac{\pi_i^I}{\pi_i^N}\right) \qquad \text{(A.15)}
\end{aligned}$$

Finally define the function $G(\rho) = \ln\left(\frac{\rho}{1-\beta}\right) + \frac{s-1}{s}\ln\left[\frac{1}{\beta}\left(\frac{\rho}{1-\beta}\right)^{\frac{s}{1-s}} - \frac{1-\beta}{\beta}\right]$ and the relative density of immigrants $\zeta_i^0 = \frac{\pi_i^I}{\pi_i^N} - \phi\sum_{i=1}^{n}\omega_i^0\frac{\pi_i^I}{\pi_i^N}$ and substitute them into expression (A.15) to derive expression (8).

B Descriptive analysis

Based on the published code from Fitzenberger and Seidlitz (2020), the Stata do-files (*d0_clean.do* and *d1_describe.do*) used to perform the descriptive analysis in Section 4 are available in a GitHub repository. This appendix discusses all the steps and procedures implemented in them.

B.1 Data cleaning and sample selection

First, all observations in the SIAB other than those from the Employment History are dropped. Those from before 2000, for individuals below age 25 or above 55, and missing employment status information are also dropped.

Second, one-time bonus payments separately reported by employers starting in 2013 are added to their annually reported daily wages. If there remain multiple overlapping spells for an individual for a given period, the one subject to social security with the highest earnings is kept, while others are dropped.

Third, observations are labeled as West German (excluding Berlin) or East German based on the state in which an individual works. Missing regional labels are then imputed. If an individual has only ever worked in one region (i.e. West or East), observations missing labels for that individual are then assigned to that region. Observations missing labels between ones with the same labels for the same individual are also interpolated, but if an individual's first or last observations are unassigned, they are not extrapolated. Observations still missing labels are then labeled as West German.

Fourth, daily wages reported at an erroneously large scale—specifically, more than 1.5 times the social security contribution threshold—are divided by 100. The remaining daily wages above this threshold are flagged and replaced with the value of the threshold itself. Observations with daily wages below the marginal part-time employment threshold are then dropped, while daily wages are deflated into 2017 terms using the Consumer Price Index.

Fifth, consecutive full- or part-time employment spells at the same employer are aggregated by taking averages of the real daily wages weighted by the durations of the spells. Fractions of the year for which the aggregated spells are valid are then calculated, natural logarithms of aggregated real wages are taken, and apprenticeships and inactive spells due to illness, maternity, or sabbaticals are dropped.

Sixth, observations are assigned to one of each of the following categories:

- Age: 25-30, 31-35, 36-40, 41-45, 46-50, or 51-55.
- Education: low, medium, or high.
- Occupation: agriculture, manufacturing in industry, technical, mining and manufacturing outside industry, service in goods trading, service in other trading, logistics, administration, healthcare, education, or other.
- Industry: agriculture and mining, manufacturing, construction, wholesale and retail trades, transportation and storage, accommodation and food service, information or communication technology and real estate or business services, finance and insurance, publication administration and defense, education, healthcare and social work, or other.

Seventh, the probability of a full-time observation from before 2012 actually being part-time according to the new and improved part-time indicator introduced in 2011 is estimated using the procedure in Fitzenberger and Seidlitz (2020). The steps of this procedure are as follows:

1. For each gender, graphically identify the upper bound for the percentile of the full-time wage distribution above which the improvement to the part-time indicator did not lead to discontinuous changes in wage growth between 2010 and 2012. The full-time weights for all observations with wages below this upper bound must then be estimated. The authors identify these upper bounds for women and men as the 80th and 25th percentiles of their respective wage distributions. Figures B1 and B2 are reproductions of the plots on whose basis they make these identifications, namely, Figures 2 and 3 of their own paper. The former shows that real wage growth for women increased sharply below the 80th percentile in 2011 before restabilizing in 2012, while the latter shows the same for men below the 25th percentile. These spikes suggest that, before 2011, there were lower-wage observations inaccurately being labeled as full-time when they actually reflected part-time work.

2. Identify the upper bound for the percentile of the total (i.e. both full- and part-time) wage distribution for each gender corresponding to the upper bound identified for the full-time wage distribution. The authors calculate this percentile to be the 88th for women and 29th for men.

3. Define the variable θ as the distance of the rank of an individual observation's wage in the total wage distribution from the upper bound.

4. Consider the following probit model:

$$P(y_{tsi} = 1|\theta_{tsi}, x_{tsi}) = \Phi(\alpha_t + \beta_t\theta_{tsi} + \gamma_t x_{tsi}) \tag{B.1}$$

   where $y_{tsi} = 1$ indicates whether the sth spell for individual i in year t is labeled as part-time; θ_{tsi} is the rank distance; x_{tsi} controls for age, age squared, education categories, occupation categories, industry categories, and states of work (including observations missing education, occupation, industry, and state data); and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Estimate equation (B.1) for each gender for each of the years between 2000 and 2012, using

Figure B1: Cumulative real wage growth for women

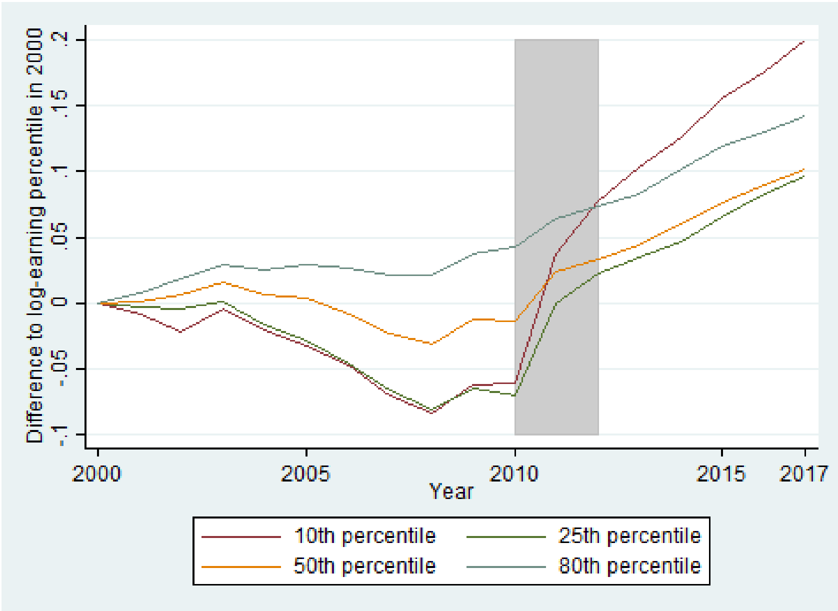
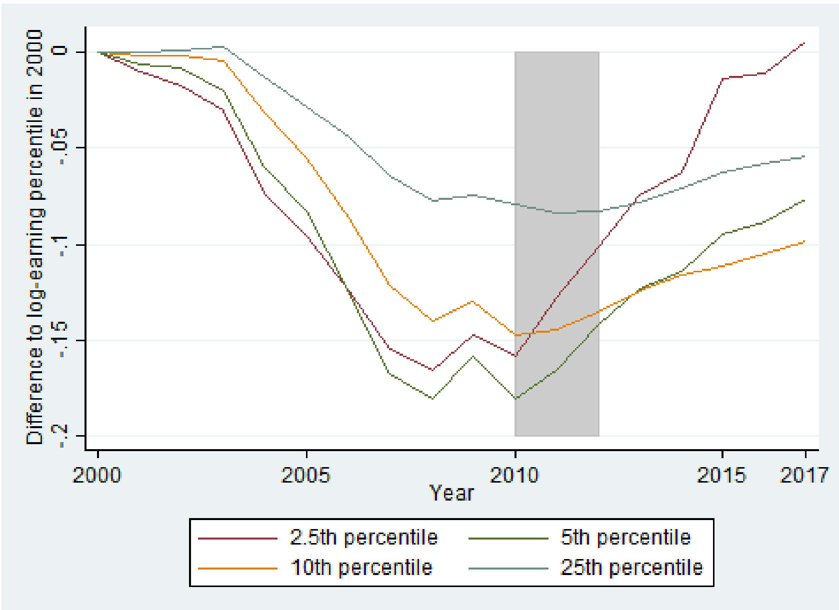


Figure B2: Cumulative real wage growth for men



Notes: Figure B1 plots cumulative real wage growth since 2000 at different percentiles of the wage distribution of weighted female full-time workers in the former states of West Germany (excluding Berlin) based on the SIAB. Figure B2 is the same plot but for men. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Nominal wages are deflated into 2017 terms using the Consumer Price Index. Weights reflect the fraction of a given year in which each full-time spell was valid.

only observations with wages below the upper bound. Tables B1 and B2 report the resulting estimates for women and men, respectively, which are reproductions of Tables 2 and 1 of the authors' paper. My estimates differ slightly from theirs because I define industry categories using the 2008 edition of the German Classification of Economic Activities, while they use the 1993 edition (Antoni et al. 2019). The fact that the estimated coefficients on most variables are greater in 2012 than in prior years for both genders is consistent with the reality that the probability of being labeled part-time is higher in 2012, after the correction to the indicator in 2011. Furthermore, the positive coefficient on θ is consistent with the expectation that, the further an observation's wage is from the upper bound of percentiles affected by this correction—that is, the lower its wage—the more likely that it is part-time.

5. Calculate the full-time weight for each pre-2012 observation with a full-time label as the ratio of their estimated probability of actually being part-time in 2012 to that of being so in the year of the observation, capped at 1. Specifically, the full-time weight is defined as:

$$\omega_{tsi}^{FT} = \min \left\{ \frac{1 - P(\widehat{y_{2012si} = 1}|\theta_{tsi}, x_{tsi})}{1 - P(\widehat{y_{tsi} = 1}|\theta_{tsi}, x_{tsi})}, 1 \right\} \tag{B.2}$$

   The intuition behind this definition is that full-time observations more likely to be part-time in 2012 than in prior years, which describes the majority of them, should be downweighted to reflect the correction to the part-time indicator. However, part-time observations are not upweighted to reflect their estimated probability of actually being full-time, since this potential source of error was not addressed by the correction. The full-time weights for observations starting in 2012 are also set to 1.

6. Calculate the final weight of each full-time observation in every year between 2000 and 2017 by multiplying its full-time weight by the fraction of a given year in which it was valid:

$$\omega_{tsi} = \omega_{tsi}^{FT} \cdot \frac{days_{tsi}}{days_t} \tag{B.3}$$

   where $days_{tsi}$ is the number of days for which the spell is valid and $days_t$ is the number of days in year t (i.e. 366 in leap years, 365 otherwise).

Table B1: Probit regressions for the part-time status of women

	(1) 2000	(2) 2005	(3) 2010	(4) 2011	(5) 2012
Rank distance from upper bound (θ)	2.832	2.558	2.395	2.888	3.108
	(148.4)	(136.7)	(135.6)	(158.5)	(162.4)
Age	0.227	0.230	0.203	0.223	0.229
	(46.4)	(48.5)	(45.7)	(51.0)	(51.9)
Age squared	-0.002	-0.002	-0.002	-0.002	-0.002
	(-40.6)	(-42.1)	(-38.5)	(-43.4)	(-44.0)
Low education	0.038	0.068	0.079	0.064	0.115
	(0.9)	(1.6)	(2.1)	(1.8)	(3.4)
Medium education	-0.015	-0.004	0.020	0.151	0.124
	(-0.4)	(-0.1)	(0.5)	(4.5)	(3.9)
High education	0.291	0.264	0.243	0.447	0.415
	(6.8)	(6.1)	(6.3)	(12.5)	(12.4)
Observations	156,514	143,407	158,221	163,708	165,381

Table B2: Probit regressions for the part-time status of men

	(1) 2000	(2) 2005	(3) 2010	(4) 2011	(5) 2012
Rank distance from upper bound (θ)	7.861	6.995	6.999	7.628	8.501
	(43.1)	(44.9)	(52.6)	(60.8)	(67.6)
Age	0.069	0.047	0.027	0.026	0.036
	(6.3)	(4.8)	(3.1)	(3.2)	(4.5)
Age squared	-0.001	-0.001	-0.000	-0.000	-0.000
	(-5.9)	(-4.2)	(-2.5)	(-2.4)	(-3.4)
Low education	-0.176	0.011	-0.065	-0.033	0.029
	(-3.0)	(0.2)	(-1.4)	(-0.8)	(0.9)
Medium education	-0.181	-0.054	-0.180	-0.059	-0.046
	(-3.2)	(-1.0)	(-4.3)	(-1.7)	(-1.5)
High education	0.410	0.429	0.289	0.370	0.375
	(6.3)	(6.8)	(5.8)	(8.5)	(9.5)
Observations	85,026	75,469	79,639	82,470	81,251

Notes: Table B1 displays the estimated coefficients (with t-statistics in parentheses) on certain variables in the probit regressions of the part-time status of women on those variables as well as occupation categories, industry categories, and states of work (including observations missing education, occupation, industry, and state data). Table B2 displays the same estimates for men. The regression is estimated for different years on the total weighted population of full- and part-time workers based on the SIAB. Included are workers who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold but below the upper bound for those affected by the improvement to the part-time indicator, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Weights reflect the fraction of a given year in which each spell was valid.

Eighth, part-time observations are dropped, such that only full-time observations remain. The wages of individuals with spells at different employers during a given year are then aggregated across spells using the weights ω_{tsi}, such that only one observation remains per individual per year. When aggregating, their weights are summed, while their maximum education and occupational skill levels are kept. Observations missing aggregated education data are dropped, those with aggregated nominal wages within 6 euros of the social security contribution threshold are labeled as censored, and the natural logarithm of aggregated real wages is again taken.

Ninth, wages are imputed for censored observations using the procedure in Gartner (2005). The steps of this procedure are as follows:

1. For each gender, estimate weighted censored regressions of aggregated log real wages on age categories, education categories, and their interactions, clustering standard errors by age-education groups.

2. Predict wages and standard errors for each observations based on the estimates of these censored regressions.

3. Define α as the deviation of these predicted wages from the relevant censoring threshold, normalized by the predicted standard errors.

4. Impute log real wages for censored observations as the sum of predicted wages and the expression $\Phi^{-1}(X[1-\Phi(\alpha)]+\Phi(\alpha))$, where $\Phi(\cdot)$ is again the cumulative distribution function of the standard normal distribution and X is a random variable uniformly distributed from 0 to 1.

Tenth and lastly, the descriptive analysis in Figures 1-2 and Tables 1-3 from Section 4 is performed using the remaining observations, which are then aggregated at the levels of the 10 former states of West Germany for every year from 2000 to 2017. At this aggregated state-year level, weighted percentiles of the native wage distribution are calculated in addition to the weighted immigrant-to-native ratios and demographic controls used in the regression analysis in Section 5.

B.2 Kernel estimates of relative density

Section 4.3 discusses how immigrants' actual and predicted wages are ranked in the native wage distribution in order to construct Figure 3. The precise steps to compute their ranks and construct this figure are as follows:

1. Based on either their actual or predicted wage, the rank of a given immigrant in the native wage distribution in a given year is the weighted percentage of natives who earn a lower actual wage during that year. Define this rank as $p \in [0, 1]$ and the associated random variable as P.

2. Compute the log odds of this rank as the function $g(p) = \ln \frac{p}{1-p}$.

3. Fit a kernel estimate of $g(p)$ across multiple years, for example, from 2011 to 2017 in Figure 3. The estimated density function is then $\widehat{f}_{g(P)}(g(p))$.

4. Retransform this estimated density function for the log odds of the rank into that of the rank itself, using the following fact from probability theory about the distributions of functions of random variables:

$$f_{g(P)}(g(p)) = f_P(p) \cdot \left| \frac{d}{dg(p)} \left[g^{-1}(g(p)) \right] \right| \tag{B.4}$$

   Since $g^{-1}(x) = \frac{e^x}{1+e^x}$ and $\frac{d}{dx}\left[g^{-1}(x)\right] = \frac{e^x}{(1+e^x)^2}$, the estimated density function for immigrants' ranks in the native wage distribution derived from equation (B.4) and plotted in Figure 3 is then:

$$\widehat{f}_P(p) = \frac{\widehat{f}_{g(P)}(g(p))}{\left| \frac{d}{dg(p)} \left[g^{-1}(g(p)) \right] \right|} = \frac{\widehat{f}_{g(P)}(g(p))}{\left| \frac{e^{g(p)}}{(1+e^{g(p)})^2} \right|} \tag{B.5}$$

5. Define the estimated density of immigrants relatives to natives at the pth percentile of the native wage distribution as $\widehat{\zeta}_p = \widehat{f}_P(p) - 1$. The density function $\widehat{f}_P(p)$ is an estimate of the first term on the right-hand side of the definition for the theoretical relative density of immigrants for each skill group $\zeta_i = \frac{\pi_i^I}{\pi_i^N} - \phi \sum_{i=1}^n \omega_i \frac{\pi_i^I}{\pi_i^N}$ stated in Section 2.2. The second term approximates 1 assuming a perfectly elastic capital supply (i.e. $\phi = 1$) and a weighted average relative density across skill groups close to 1 (i.e. $\sum_{i=1}^n \omega_i \frac{\pi_i^I}{\pi_i^N} = 1$).

C Regression analysis

The Stata do-file (*d2_regress.do*) used to perform the regression analysis in Section 5 is available in the same GitHub repository as before. This appendix presents additional results from these regressions.

C.1 Residual serial correlation tests

Table C1: Arellano-Bond tests for first- and second-order serial correlation

	First-Order	Second-Order
5th percentile	1.207	-0.355
	(0.227)	(0.722)
10th percentile	-0.724	-0.974
	(0.469)	(0.330)
15th percentile	-1.151	-0.804
	(0.250)	(0.421)
20th percentile	-1.157	−1.085
	(0.247)	(0.278)
25th percentile	0.534	-0.751
	(0.593)	(0.453)
30th percentile	0.237	0.458
	(0.812)	(0.647)
35th percentile	-1.153	0.768
	(0.249)	(0.443)
40th percentile	-1.283	-0.883
	(0.199)	(0.377)
45th percentile	0.071	-1.387
	(0.943)	(0.165)
50th percentile	0.017	-1.637
	(0.987)	(0.102)
55th percentile	0.425	-1.025
	(0.671)	(0.305)
60th percentile	-1.583	0.074
	(0.113)	(0.941)
65th percentile	-1.728	-0.194
	(0.084)	(0.846)
70th percentile	-0.207	0.462
	(0.836)	(0.644)
75th percentile	0.305	-0.341
	(0.760)	(0.733)
80th percentile	0.617	0.421
	(0.537)	(0.674)
85th percentile	0.941	0.771
	(0.347)	(0.441)

Notes: Displayed are the t-statistics and associated p-values (in parentheses) of Arellano-Bond tests for first- and second-order serial correlation in the estimated residuals from first-difference IV regressions of different percentiles of native wages on the immigrant-to-native ratio, year fixed effects, and demographic controls for the 2011-2017 period. The instrumental variable for the annual change in the ratio used in the IV regressions is its four-year lagged level.

C.2 Summary statistics

Table C2: Summary of all variables

	Mean	S.D.	Min	Max
Natives' log real wage percentiles				
5th	3.959	0.067	3.827	4.096
10th	4.147	0.064	4.009	4.282
15th	4.268	0.061	4.130	4.404
20th	4.359	0.063	4.218	4.491
25th	4.433	0.063	4.290	4.559
30th	4.496	0.064	4.354	4.622
35th	4.554	0.065	4.407	4.677
40th	4.609	0.067	4.461	4.728
45th	4.661	0.069	4.509	4.780
50th	4.714	0.073	4.557	4.836
55th	4.768	0.077	4.601	4.900
60th	4.824	0.082	4.649	4.963
65th	4.883	0.087	4.705	5.036
70th	4.947	0.091	4.764	5.105
75th	5.018	0.094	4.831	5.182
80th	5.098	0.095	4.905	5.253
85th	5.189	0.092	5.020	5.359
Immigrant-to-native ratio	0.092	0.032	0.036	0.166
Annual change in the immigrant-to-native ratio	0.004	0.003	-0.004	0.011
Average native age	41.077	0.388	40.218	41.749
Average immigrant age	39.744	0.582	38.553	41.399
Log high- to low-educated native ratio	1.766	0.354	1.192	2.346
Log medium- to low-educated native ratio	3.046	0.143	2.684	3.301

Notes: Displayed are the mean, standard deviation, minimum, and maximum of all variables across regions and time used in the regressions of native wage percentiles on the immigrant-to-native ratio, year fixed effects, and demographic controls. All variables are constructed using the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time label being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020). Education is based on a variable imputed according to the procedure in Thomsen, Ludsteck, and Schmucker (2018): low for completion of secondary school, medium for upper secondary school or vocational training, and high for university.

Table C3: Summary of the change in the immigrant-to-native ratio

	In percentage points			
	Mean	S.D.	Min	Max
2000-2001	-0.032	0.154	-0.177	0.356
2001-2002	-0.089	0.184	-0.364	0.259
2002-2003	-0.288	0.190	-0.581	0.090
2003-2004	-0.199	0.340	-0.901	0.411
2004-2005	-0.162	0.181	-0.401	0.149
2005-2006	0.141	0.210	-0.174	0.499
2006-2007	0.061	0.214	-0.337	0.448
2007-2008	0.126	0.169	-0.092	0.460
2008-2009	-0.264	0.258	-0.847	0.043
2009-2010	0.163	0.204	-0.209	0.508
2010-2011	0.495	0.179	0.210	0.800
2011-2012	0.450	0.189	0.220	0.793
2012-2013	0.285	0.270	-0.086	0.772
2013-2014	0.367	0.368	-0.271	0.824
2014-2015	0.495	0.336	-0.063	0.926
2015-2016	0.330	0.216	-0.047	0.722
2016-2017	0.541	0.399	-0.369	1.052
Average 2000-2010	-0.054			
Average 2010-2017	0.423			
2010-2017	2.964	1.362	1.450	5.513

Notes: Displayed are the mean, standard deviation, minimum, and maximum of the annual percentage-point change in the immigrant-to-native ratio across all regions between each pair of years. The second- and third-to-last rows report means across regions and time for the relevant periods. The last row reports the mean, standard deviation, minimum, and maximum for the change in the immigrant-to-native ratio across all regions between 2010 and 2017, the period which covers the recent immigration waves. The ratio is constructed using the SIAB, based on weighted full-time workers in the former states of West Germany (excluding Berlin) who are between 25- and 55-years-old, earn wages above the marginal part-time employment threshold, and are neither apprentices nor inactive due to illness, maternity leave, or sabbaticals. Immigrants are defined by their lack of German nationality. Weights reflect both the fraction of a given year in which a worker was employed full-time and the probability of their full-time indicator being accurate, estimated according to the procedure in Fitzenberger and Seidlitz (2020).