CESifo WORKING PAPERS

10344
2023
March 2023

Immigration and the Slope of the Labor Demand Curve: The Role of Firm Heterogeneity in a Model of Regional Labor Markets

Andrea Ariu, Tobias Müller, Tuan Nguyen

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10344*

Immigration and the Slope of the Labor Demand Curve: The Role of Firm Heterogeneity in a Model of Regional Labor Markets

Abstract

In this paper, we provide new explanations for the puzzling findings in the literature that migrants do not decrease natives' wages, and that skilled immigration can actually increase them. We develop a model with regional labor markets and heterogeneous firms in which workers of different skill levels are imperfect substitutes, but for a given skill level, natives and migrants are perfect substitutes within a firm. In this setting, a skilled labor supply shock due to immigration has two consequences. First, it induces skill-intensive firms and skill-abundant regions to expand. These across-firm and across-region reallocations reduce the within-firm and within-region substitution between skilled and unskilled workers, thus limiting relative wage adjustments. Second, the average native's wage can be partially sheltered from the negative effect of immigration depending on the geographical settlement patterns of immigrants. Both mechanisms make natives and migrants appear as imperfect substitutes at the aggregate level. Quantitatively, our simulations show that the negative impact of immigration on natives' wage is halved when the across-firm and across-region reallocation mechanisms are at work. Finally, both theory and simulations show that when these mechanisms are coupled with human-capital externalities that are skill-neutral at the firm level but skill-biased on aggregate, skilled immigration can increase absolute and relative skilled wages. Therefore, firm heterogeneity, local labor markets, and human-capital externalities are crucial for understanding the impact of immigration on natives' wages.

JEL-Codes: F220, J610, J310.

Keywords: immigration, firm heterogeneity, wages.

Andrea Ariu
Department of Economics, Management and Quantitative Methods
University of Milan / Italy
andrea.ariu@unimi.it

Tobias Müller
Institute of Economics and Econometrics,
GSEM, University of Geneva / Switzerland
tobias.mueller@unige.ch

Tuan Nguyen
Institute of Economics and Econometrics,
GSEM, University of Geneva / Switzerland
tuan.nguyen@unige.ch

March 23, 2023
This research was supported by the National Center of Competence in Research NCCR - On The Move funded by the Swiss National Science Foundation. We thank Giovanni Facchini, Albrecht Glitz, Jessica Hug, Marcelo Olarreaga, Michele Pellizzari, Farid Toubal, and the participants at various seminars and conferences for helpful suggestions.

1 Introduction

Standard economic theory predicts that immigration puts pressure on wages of native workers. However, with few notable exceptions (e.g., Borjas, 2003; Borjas and Katz, 2007; Monras, 2018), the majority of papers analyzing this phenomenon fail to identify a substantial negative effect (e.g., Card, 2001, 2012; Manacorda et al., 2012; Ottaviano and Peri, 2012) or they even find a positive one (e.g., Peri et al., 2015; Beerli et al., 2021). To explain this puzzle, most contributions focus on the features that make immigrants different, and thus imperfect substitutes, with respect to natives.[1] At the same time, this debate has adopted cities, regions, or countries as the unit of observation, mostly ignoring the role of firms in this process, despite the increasing availability of employer-employee data.[2]

This paper takes the opposite approach: we study the puzzling fact that immigration does not seem to affect natives' wages: we assume that migrants and natives are perfectly substitutable at the firm level for a given skill, and we build a model that describes how firms and wages react when confronted with an increase in the number of foreign workers. We find that two forces can dampen the impact of immigration on wages. First, if firms are heterogeneous in terms of the skill composition of their workforce, an increase in the supply of workers in a particular category (e.g., skilled) favors firms using them intensively. Akin to Rybczynski effects, a large share of the increase in the supply of labor would be absorbed by expanding firms. Second, in the absence of perfect labor mobility, regions receiving more immigrants would see their production expand more, causing a redistribution of national output. These across-firm and across-region reallocations limit the extent of within-firm and within-region adjustments, thereby weakening the wage effects of immigration.

The latter mechanism operates even more to the advantage of natives when migrants and natives are not homogeneously distributed across regions. The (partial) segregation between native and migrant workers of the same skill partially shelters natives from the negative wage impact of immigration, while migrants are more exposed. As a result, even if natives and migrants are perfect substitutes at the micro level, on aggregate they can appear as being imperfect substitutes.

Using the model to perform counterfactual exercises, we find that the impact of immigrants on natives' wages would be twice as important without these mechanisms. Therefore, they play an important role in the labor-market relation between natives and immigrants. Moreover, we show that when adding human-capital externalities at the regional level that are neutral to the firm but skill-biased in the aggregate, the wage effect of immigration can actually even turn positive. In our model, the aggregate skill bias can be sufficiently strong to imply that skilled immigration increases skilled wages more than unskilled wages. This appears to be the empirically relevant case for Switzerland, as shown by Beerli et al. (2021) and confirmed by our estimations.

These results provide a new perspective to explain the puzzling finding that immigrants do

[1] For example, Peri and Sparber (2009) focus on different task specialization; Dustmann et al. (2012) on the educational downgrading of immigrants who, given the same educational level, accept worse jobs than natives; and Llull (2017) on the occupational choices of natives facing immigration.

[2] The most notable exception is Dustmann and Glitz (2015).

not seem to negatively affect natives' wages, and they highlight the importance of having firm heterogeneity as the focus of analysis to understand the labor-market outcomes of migration. In addition, our model also provides a novel microfoundation for the empirical finding that skilled immigration can increase skilled wages more than unskilled wages.

Our analysis is organized in three sections. First, we use the Swiss Earnings Structure Survey (SESS) to show that heterogeneity across firms and regions is a pervasive feature of our data. Specifically, firms differ in terms of skill composition of employment (also within sectors). This means that they use different technologies (e.g., share of skilled workers) to produce similar output. Thus, an increase in the supply of a particular category of workers can have different effects depending on their utilization across firms, even within the same sector. At the same time, the presence of foreign workers varies across regions, and new immigrants tend to settle in those with large diasporas. Therefore, an increase in the supply of migrants can have heterogeneous effects also across regions. Finally, we show that natives do not move in response to migration shocks. This provides the basis for the assumptions of segmented regional labor markets.

Second, we build a model of labor markets where we incorporate the data features just described. Each firm is a monopolist in the production of a differentiated intermediate product that is used for the production of a final product. Firms differ in terms of technology in the sense that they employ different proportions of skilled and unskilled labor. Moreover, we assume that labor markets are segmented and residents cannot move across regions. We use this model to study the impact of an increase in labor supply of skilled workers due to immigration.

The model delivers two main mechanisms. One, within the same local labor market, firm heterogeneity makes an increase in the supply of skilled workers cause a reallocation from unskilled intensive firms to skilled intensive ones. This creates an across-firm reallocation process that limits the extent of relative wage adjustments. Two, if migrants locate unevenly across regions, the regions receiving a greater number would increase their production more than those receiving less. This would cause an across-region reallocation of national output that would decrease wages predominantly in regions with high shares of immigrants and shelter natives from the negative wage impact of immigration.

Both mechanisms are magnified in the case of highly substitutable products. This is because if the intermediate products are close substitutes, firms using skilled workers intensively and those regions receiving more of them get to produce an even larger share of output. To complete the discussion, we analyze how relaxing the assumption of worker immobility across regions would change the predictions of our model.

Then, we augment the model by adding human-capital externalities at the regional level, as in Moretti (2004)—we analyze the role of firm heterogeneity in this context. We assume that these externalities are skill-neutral at the firm level and show that their effect can become skill-biased in the aggregate if the impact of the externality on productivity is heterogeneous at the firm level. As a result, the effect of immigration on natives' wages can actually turn positive as long as skilled intensive firms benefit more from a larger pool of skilled workers. Specifically, if the correlation between the skill share at the firm level and the sensitivity of each firm to the externality is positive and sufficiently large, skilled immigration can even lead to an increase in the *relative* wages of skilled workers.

Third, we perform counterfactual exercises to understand the force of the mechanisms arising from our model. The employer-employee data from the SESS allow us to estimate the elasticity of substitution between skilled and unskilled workers at the firm level and to calibrate all the other parameters, with one exception: the elasticity of substitution across products, which we take from Egger et al. (2012). To address the endogeneity issues arising from unobserved demand or supply shocks that could bias our estimates of the firm-level substitution elasticity between skilled and unskilled labor, we use a Card (2001) shift-share instrument in which the share is represented by the 1980 settlement of immigrants in each labor market area by country of origin and the shift by the national arrival of immigrants during the 2004–2012 period by origin and education. This IV strategy performs well, allowing us to provide the literature with an elasticity that, with the exception of Bøler (2015), has been estimated only at the aggregate level.

We then perform two counterfactual exercises. In the first, we reallocate the observed increase in labor supply due to immigration during the 2004–2010 period at the national level based on the distribution of immigrants of the same nationality across regions following the approach of Card (2001). In the second, we increase the number of cross-border workers (CBW) across labor markets, depending on their time to travel to the nearest border crossing, thus mimicking the labor supply shock brought by the implementation of the Agreement on the Free Movement of Persons (AFMP) between Switzerland and the European Union, described in Losa et al. (2012), Beerli et al. (2021), and Ariu (2022).

The first shock is stronger but more evenly distributed across regions. The second is weaker but more regionally clustered, allowing us to evaluate in particular the across-region reallocations. The results of these exercises show that the presence of heterogeneous firms and regional labor markets reduces—by half—the negative impact of an increase in the relative supply of skilled to unskilled labor due to immigration. The first mechanism accounts for most of the attenuation effect, while the second plays a more marginal role. Adding the human-capital externalities into our simulations, we show that the effect of an increase in the number of skilled migrants can actually be positive and sizeable for natives' wages.

This paper is related to the lively debate on the wage effects of immigration. Few papers find a negative effect and thus evidence of substitutability between natives and migrants (e.g., Borjas, 2003; Borjas and Katz, 2007; Monras, 2018), whereas the majority of papers find an imperfect substitutability or complementarity (e.g., Card, 2009; D'Amuri et al., 2010; Ottaviano and Peri, 2012; Manacorda et al., 2012; Beerli et al., 2021). Rather than taking a stand on how to correctly estimate this relation, our paper provides simple mechanisms that, even in the extreme case of perfect substitutability at the micro level, can explain why they can appear at aggregate level as imperfect substitutes or even complements.

Most of the existing contributions explain the substitutability puzzle by arguing that immigrants are hardly comparable with respect to natives, even within the same education group. For example, natives and migrants select into different occupations following their comparative advantage in communication tasks (Peri and Sparber, 2009); migrants tend to accept jobs for which they are overqualified when they first arrive (Dustmann et al., 2012); or natives adapt their occupational choices when facing immigration (Llull, 2017). In this paper, we take the opposite perspective, analyzing mechanisms that can make immigrants look like imperfect substitutes (or complements) in the aggregate even when considering them as perfect substitutes at the micro level.

In this sense, the closest papers are Burstein et al. (2020) and Dustmann and Glitz (2015). The first relates the apparent imperfect substitutability between natives and migrants to the capacity of the worker's occupation to adjust to the immigration shock by exporting. Besides focusing on a different mechanism, we also take a different perspective by putting firms and their heterogeneity at the center of analysis, by assuming that natives and migrants are perfect substitutes at the micro level, and by showing that the low substitutability arises also when the extra output caused by immigration cannot be absorbed by international markets. The second paper empirically analyzes how labor markets react to immigration through changes in wages, firm expansion, and within-firm adjustments using German data. Our paper provides a theoretical framework that puts together both the firm and regional dimensions to identify and quantify the role of firm and regional heterogeneity in explaining the impact of immigration on natives' wages.

The literature on the economics of migration only recently started to analyze the impact of migration on the economy with the firm as the focus of analysis. Beerli et al. (2021) use the AFMP between Switzerland and the EU to show that firms more exposed to the supply shock of immigrants increased their employment, skill intensity, sales, and productivity following its implementation. Kerr and Lincoln (2010), Ghosh et al. (2014), and Mayda et al. (2020) reach similar results using the 2004 U.S. cap on H-1B visas as exogenous sources of variation. Hornung (2014) uses the expulsion of Huguenots from France in 1685 and their subsequent random reallocation in Prussia to show that they increased the productivity of Prussian firms in the textile sector. Mitaritonna et al. (2017) show a positive effect of immigrants on total factor productivity of firms in France. Cristelli and Lissoni (2020) and Gray et al. (2020) focus on the effect of immigration on innovation, and Kugler and Rapoport (2007), Javorcik et al. (2011), and Burchardi et al. (2019) on FDI activity. With respect to these contributions, our paper offers further evidence that the firm dimension is key to understanding the effects of migration on the economy and the relation between native and migrant workers.

More broadly, our paper highlights the importance of heterogeneity in understanding aggregate outcomes. Examples of this growing literature include Melitz (2003), who analyzes the importance of productivity heterogeneity across firms within the same sector to understand the consequences of trade liberalization on firms' export activities and aggregate productivity; Oberfield and Raval (2021), who disaggregate the capital-labor elasticity of substitution into the across- and within-plant components to understand the decrease in labor's share of income observed in the United States; and Baqaee and Farhi (2019), who analyze the macroeconomic impact of microeconomic productivity shocks. Our paper shows that heterogeneity across firms is key also to understanding the impact of immigration on natives' wages. Specifically, heterogeneity limits within-firm adjustments in favor of across-firm adjustments, thus limiting the need for firms to substitute natives with migrants.

There are only a few papers that incorporate firm heterogeneity in the context of migration. Di Giovanni et al. (2015) show in a cross-country setting that the welfare consequences of migration can depend on productivity differences across firms. Using German employer-employee data, Brinatti and Morales (2021) provide evidence that big firms employ a higher share of immigrants than small firms, and accounting for this heterogeneity is crucial for correctly estimating the welfare effects of immigration. Emami Namini et al. (2015) interpret firm heterogeneity in terms of factor intensity, and find that an increase in skilled competition due to migration can nullify the positive effects of trade liberalization on TFP and real income.

The rest of the paper is organized as follows. Section 2 describes the data and presents three main stylized facts. Section 3 outlines the theoretical model. Section 4 provides a quantification of the mechanisms arising from our model. Section 5 concludes.

2 Data and Stylized Facts

In this section, we describe the data used for the analysis and outline three stylized facts that motivate the structure of the theoretical model.

2.1 Data

Our main source of information is the Swiss Earnings Structure Survey (SESS). This large, cross-section survey has been carried out every two years on a representative sample of workers and firms since 1994 by the Swiss Federal Statistical Office (FSO). The advantage of this dataset is that, besides the usual individual information about workers (e.g., wage, education, and occupation), it also includes details of the firm in which the worker is employed, such as the industry, the size of the firm, and its location. For our analysis, we limit our sample to employees working in the private sector (excluding agriculture), aged 18 to 65, over the 1994–2014 period, and for whom we have complete information.

We aggregate the nine education categories of the survey into two broad categories: skilled (tertiary education) and unskilled (primary and secondary education). We use a full-time equivalent wage rate. The gross monthly earnings are standardized by the FSO at 40 hours per week, and we drop the top and bottom 0.5% of the wage distribution in each year. The survey does not provide the country of birth of workers, only their work permit (e.g., Swiss citizen, settlement permit, cross-border worker, or short-term residence permit). Thus, we will use this information to define Swiss citizens as native workers and those who do not hold Swiss citizenship (i.e., the foreigners) as immigrant workers.

We aggregate the individual information at the firm-year level. Precisely, for every firm and year, we have information on the number of full-time equivalent workers by educational groups and their respective average wage. We also have information on the size of the firm measured as the total number of employees, its industry at the NOGA 2008 two-digit level (which corresponds broadly to a NACE two-digit classification), and the MS-region in which the firm is located. MS stands for "mobilité spatiale" or spatial mobility areas and we count 106 of them. These have been defined by the FSO based on observed commuting patterns and economic activities. As such, they can be seen as local labor markets.[3]

Our final sample includes about 8 million observations, thus we have a large number of firms for every sector and region, which allows us to exploit the different sources of variation present in our dataset. Unfortunately, the firm identifier changes every year, therefore, we are not able to track firms over time. Moreover, we do not have information about capital at the firm level. Below, we provide three stylized facts that support and motivate the assumptions of the theoretical model.

2.2 Stylized Facts

In this subsection, we present three stylized facts, based on the data just described, that support the assumptions and the structure of the model.

- ***Fact 1:*** *The skill composition of employment differs both across and within sectors.*

In Switzerland, about a fifth of workers are skilled. Looking at their distribution across sectors, we observe that their importance varies quite substantially. Column 1 of Table 1 shows that their share can vary from 6.9% of the mining sector to 58.7% of the R&D sector. Clearly, the use of skilled workers in production varies depending on the products. Moreover, the number of firms with at least one skilled worker is also widely heterogeneous, varying from 26.3% of the retail sector to 90.8% of the R&D sector. Therefore, the use and importance of skilled workers across industries varies widely: for example, the hotel and restaurant sector and the mining sector require low cognitive activities, so only a small fraction of workers in these industries needs to be skilled and fewer firms employ skilled workers. Conversely, the tasks performed in the R&D sector have a high cognitive content, which implies a large share of skilled workers and a large share of firms employing high-skilled workers.

The number and importance of skilled workers are also quite uneven across firms *within* sectors. Column 2 of Table 1 highlights that not all firms have skilled workers within the same sector. This means that even within sectors, firms differ in their need for skilled workers. This can be explained by the different technologies that firms can adopt to produce similar products. For example, the same garment can be assembled by a machine or handmade by an unskilled worker. For example, in the education sector, child care workers are frequently unskilled, and, in the R&D sector, some firms provide unskilled R&D support services (such as data entry and preparation or testing of basic materials) that employ almost exclusively unskilled workers. This is further highlighted by

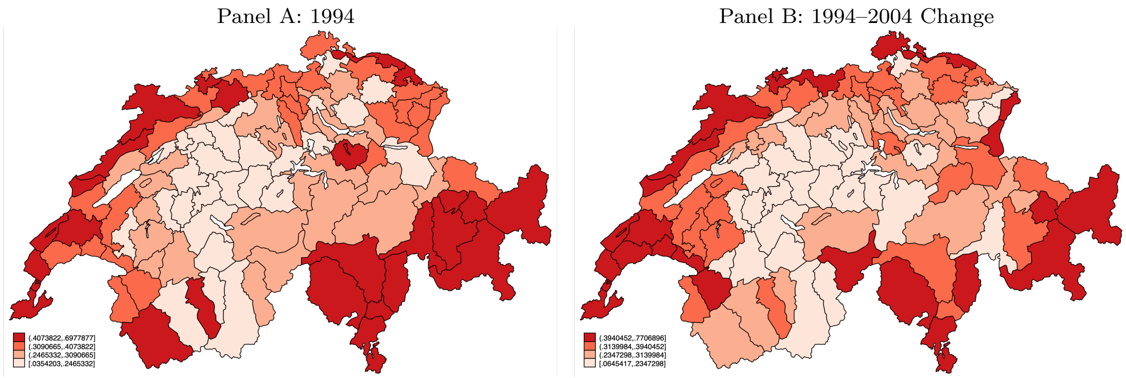
[3] The survey contains the postal code of the place of work, but for reasons of confidentiality, the FSO provides that information only at the broader MS-region level. We treat firms having multiple plants in different regions as being different firms.

Table 1: Share of High-Skilled Workers by Sector, 1994–2014

Sector Noga (2-digits)	(1) Share of Skilled	(2) Share of Firms with Skilled	(3) Share of Skilled per Firm	(4) Standard Deviation of Share of Skilled per Firm
Mining and Quarrying (5-9)	6.9	33.8	6.4	14.5
Food and Beverage (10-11)	11.9	47.9	13.4	23.4
Tobacco (12)	21.3	70.4	16.1	21.1
Textile (13-15)	10.1	35.4	10.3	22.0
Wood and Paper (16-18)	10.5	43.9	11.1	20.7
Chemical (19-20)	19.5	58.7	19.7	26.2
Pharmaceutical (21)	33.6	83.3	34.2	29.6
Plastic and Rubber (22-23)	11.6	47.6	10.8	19.2
Metals (24-25)	12.4	53.4	12.3	19.7
Computer (26)	29.3	69.9	28.9	31.1
Electrical Machinery (27)	29.6	54.9	19.3	26.2
Other Machinery (28)	24.0	62.9	21.6	25.6
Vehicles (29-30)	20.8	59.8	19.4	26.1
Other Manufacturing (31-33)	16.8	44.4	15.3	25.5
Produc. and Distrib. of Energy (35)	28.2	74.0	25.1	25.9
Water and Garbage Manag. (36-39)	8.7	34.2	9.0	19.4
Construction (41-43)	10.7	44.7	11.3	20.0
Wholesale (45-46)	18.7	45.8	17.0	27.0
Retail (47)	8.9	26.3	8.2	19.9
Transport (49-52)	9.1	38.0	10.1	21.1
Postal and Courier Activities (53)	8.6	28.0	8.6	21.5
Hotels and Restaurants (55-56)	6.9	30.6	6.9	16.8
Audiovisual (58-60)	37.9	67.4	38.5	37.6
Telecommunication (61)	35.0	61.3	29.0	33.3
IT (62-63)	52.4	79.7	59.3	36.8
Finance (64,66)	33.6	69.7	35.4	33.8
Insurance (65)	29.7	53.7	19.5	26.9
Real Estate (68)	22.7	53.4	25.0	31.3
Legal, Architectural and Engineering (69-71)	50.9	77.8	49.9	35.4
R&D (72)	58.7	90.8	69.0	32.3
Other Technical Services (73-75)	30.8	58.7	34.3	36.8
Administrative Services (77,79-82)	11.1	32.6	12.1	24.0
Employment Services (78)	18.5	35.6	25.0	36.3
Education (85)	51.3	75.4	47.0	37.6
Health (86-88)	25.4	63.0	27.5	32.0
Art, Culture (90-93)	26.0	50.5	24.2	33.0
Associations and Personal Services (94-95)	42.1	63.8	43.0	38.1
Other Services (96)	6.5	17.9	7.5	21.2

Notes: This table shows the average for the 1994–2014 period for the share of skilled workers, the share of firms with at least one skilled worker, the share of skilled workers per firm and its standard deviation by Noga two-digit sector. Data source: SESS.

Figure 1: Share of Foreign Workers by MS-Region



Notes: Panel A depicts the share of foreign workers by MS-region in 1994. Panel B shows the change in the same share between 1994 and 2014. Data source: SESS

the fact that the average share of skilled workers per firm (column 3) is different from the sectoral mean (column 1), and by the fact that the standard deviation in the firm share of skilled workers (column 4) in the same sector tends to be quite high. This evidence suggests that the impact of a supply shock of skilled workers (e.g., coming from abroad) could vary not only across sectors but also across firms within the same sector. Such a shock would probably affect universities or research laboratories but not kindergartens and unskilled service R&D producers.

- ***Fact 2:*** *Foreign workers spread unevenly across regions.*

The geographical distribution of immigrants represents another important dimension to be taken into account to understand their impact on natives. Figure 1 represents the share of foreign workers across our 106 MS-regions in 1994 (Panel A). Their share with respect to the overall number of workers varies from 3.5% to 69.7%. Not surprisingly, regions with the highest share of foreign workers are located close to the border. This is mostly because cross-border workers (i.e., those who work in Switzerland but live in one of the border countries) have to return to their home country after work and are disinclined to travel long distances.

This geographical heterogeneity is important, because new foreign workers tend to settle in regions where there is already a substantial diaspora (e.g., Card, 2001; Beine et al., 2011). Indeed, Panel B shows that the regions that increased their share of foreign workers are located closer to the border and already had a high share of foreign workers. This means that a supply increase of foreign workers from foreign countries can have heterogeneous effects across regions depending on where they settle.

- ***Fact 3:*** *Immigration does not influence natives' location choices.*

If natives tend to relocate to other regions following an immigration shock, carrying out an analysis at the local level could potentially bias the estimation of the wage effects on natives (e.g., Borjas, 2006; Monras, 2018). In this case, the aggregate (i.e., national) level should be the appropriate one for the analysis. Instead, if natives do not react to the immigration shock, local labor markets are the appropriate unit of analysis (e.g., Peri and Sparber, 2011). In the Swiss context, Beerli et al. (2021) show that foreign workers did not affect natives' labor mobility across municipalities following the implementation of the AFMP.

To provide further evidence in the Swiss context, we use new data from the Swiss Census from 1990 and 2010 to assess whether municipalities that experienced an increase in the number and share of immigrants saw their native population shrink. To address the usual endogeneity concerns, we exploit the empirical strategy proposed by Beerli et al. (2021). Specifically, we use the variation in the intensity at which Swiss municipalities were affected by the implementation of the AFMP. This agreement dropped all restrictions for hiring EU workers and led to an increase in the number and share of foreign workers that varied depending on the distance from the border, mostly because cross-border workers are disinclined to commute long hours to get to work.[4]

To implement this empirical strategy, we regress population changes between 1990 and 2010 for each municipality m using Swiss Census data on dummies identifying the driving time to the border, D1_m for within-15 minutes, and D2_m for 15 to 30 minutes. The results of this regression can assess whether municipalities within 15 minutes of the border or 15 to 30 minutes from the border experienced differential changes in population growth with respect to municipalities beyond 30 minutes from the border, distinguishing natives from immigrants. This exercise is instructive for two reasons. First, we can test whether the flow of migrants is due to unobserved demand factors by checking whether immigrant flows experienced the same dynamics as natives. In this case, we should observe that the number of natives and migrants move in tandem with respect to the shock. Second, we can test whether natives are displaced by immigrants by checking whether their number decreased following the implementation of the AFMP.

Table 2 shows that there is no significant differential increase in the population across municipalities depending on distance from the border. Moreover, by distinguishing between Swiss nationals and foreigners, we observe that the differential increase of immigrants in treated municipalities is not followed by changes in native population. Therefore, migrant flows should not be driven by unobserved factors, and they do not cause movements across municipalities for the native population. Unfortunately, the same exercise cannot be performed with the SESS data, because each MS-region can contain municipalities that are in both the central and the border regions and with different time distances with respect to the border, thus making it difficult to assign them to treatment and control.

These three stylized facts highlight the importance of firm and geographical heterogeneity in analyzing the immigration consequences. Different firms and regions could be differently impacted by a worker supply shock. This depends on the skill level of the incoming flow and the degree of heterogeneity in the skill intensity of the production of the firm. Moreover, it depends on where foreign workers decide to relocate and on the potential reaction of natives to the shock. We also find some evidence that the immigration shock does not displace natives. This result combined

[4] Please refer to Beerli et al. (2021) and Ariu (2022) for a more extensive description of the agreement and its consequences.

Table 2: Population Dynamics, 1990–2010

	(1) Δ Total	(2) Δ Swiss	(3) Δ Foreigners
$D1_m$	0.014	0.004	0.017^b
	(0.016)	(0.013)	(0.008)
$D2_m$	-0.015	-0.010	-0.005
	(0.011)	(0.009)	(0.005)
Observations	1,731	1,731	1,731
R^2	0.002	0.001	0.006

Notes: Δ Total, Δ Swiss, Δ Foreigners indicate, respectively, the change at the municipality level between 1990 and 2010 in the resident population, Swiss population, and immigrant population. $D1_m$ indicates municipalities within 15 minutes from the border crossing, and $D2_m$ indicates municipalities between 15 and 30 minutes from the border crossing. Robust standard errors are in parentheses. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1. Data Source: Swiss Census.

with the findings of Beerli et al. (2021) suggest that resident workers are not mobile across regions and that the local labor market is the right unit of analysis for the investigation of the impact of immigration in Switzerland.

3 A Model of Heterogeneous Firms and Regional Labor Markets

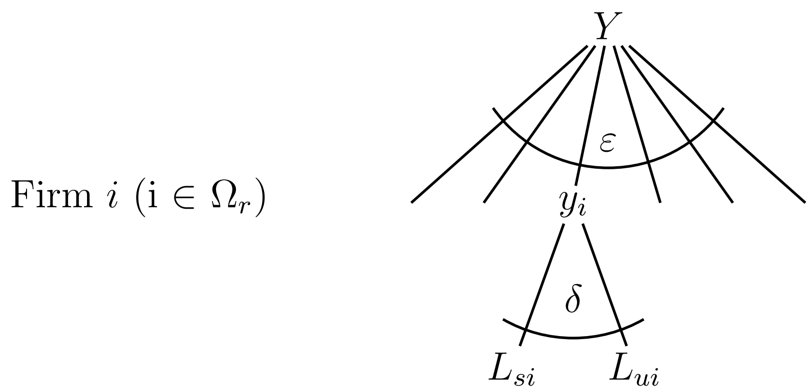
In this section, we use a simple model with heterogeneous firms and regional labor markets to analyze the impact of immigration, interpreted as a regional labor supply shock, on natives' wages. Based on the stylized facts, we assume that firms are heterogeneous in their composition of employment (i.e., the ratio of skilled to unskilled differs across firms). For a given skill level, natives and migrants are perfect substitutes in production within a firm, but workers of different skill levels are imperfect substitutes. Moreover, we assume that labor markets are segmented and residents cannot move across regions. We start by describing the assumptions and equations of the model. Then, we analyze the consequences of a shock to labor supply due to immigration on relative and absolute wages. Finally, we evaluate the role of firm and regional heterogeneity, and labor mobility in this process, and we analyze the consequences of assuming human-capital externalities at the regional level.

3.1 The Benchmark Model

There is a representative firm producing a single final good Y, which is sold in a perfectly competitive market at the national level, as depicted by Figure 2. This firm uses intermediate inputs y_i produced by a large number of firms i that operate in the R regional labor markets. The technology of the final-good firm is CES:

$$Y = \left(\sum_{i \in \Omega} \beta_i y_i^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{1}$$

Figure 2: Structure of the Aggregate Production Function



Note: This figure depicts the aggregate production function.

where ε is the elasticity of substitution between the products of different firms and Ω is the set of firms in the country. This set is partitioned into subsets Ω_r ($r = 1, \ldots, R$), where Ω_r includes the firms operating in region r.

The production function of firm i is given by a CES function with constant returns to scale. To keep the model as simple as possible, we assume that the only factor of production is labor and that in each region a fixed number of firms operate under monopolistic competition.[5]

$$y_i = A_i \left(b_{si} L_{si}^{\frac{\delta-1}{\delta}} + b_{ui} L_{ui}^{\frac{\delta-1}{\delta}} \right)^{\frac{\delta}{\delta-1}} \tag{2}$$

where δ is the elasticity of substitution between skilled and unskilled labor, which is common for all firms. Firms are, however, heterogeneous with respect to the relative efficiency of skilled and unskilled labor in production (i.e., parameters b_{si} and b_{ui} vary across firms). Total factor productivity, A_i, is exogenous in the benchmark version of the model.[6] Following the standard assumption of monopolistic competition, we assume that firm i is small relative to the sector and profit maximization leads to the condition that its price, p_i, is a fixed markup over marginal cost, c_i.[7] By Shephard's lemma, firm i's conditional labor-demand functions are given by

$$L_{si} = A_i^{\delta-1} b_{si}^{\delta} \left(\frac{w_{sr}}{c_i} \right)^{-\delta} y_i, \qquad L_{ui} = A_i^{\delta-1} b_{ui}^{\delta} \left(\frac{w_{ur}}{c_i} \right)^{-\delta} y_i, \qquad (i \in \Omega_r) \tag{3}$$

[5]An equivalent assumption to the absence of capital would be that the supply of capital is perfectly elastic at a given rental rate and that the share of capital in a firm's cost is constant across firms.

[6]In section 4.1.1, we discuss the possibility that TFP depends on the ratio of skilled to unskilled labor in the region where the firm operates, due to external effects.

[7]i.e., $p_i = \left(\frac{\varepsilon}{\varepsilon-1}\right) c_i$ and $c_i = \left(b_{si}^{\delta} w_{sr}^{1-\delta} + b_{ui}^{\delta} w_{ur}^{1-\delta}\right)^{\frac{1}{1-\delta}}$, where w_{sr} and w_{ur} are, respectively, skilled and unskilled wage rates in region r (where firm i is located, $i \in \Omega_r$). Strictly speaking, the production function (1) should be written as $Y = \left(\int_{i\in\Omega} \beta(i) y(i)^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}$, where i is a continuous index variable of firms. Then firm i's perceived demand function is given by $y_i = k_i p_i^{-\varepsilon}$, since a variation of firm i's price has no impact on the price P at the aggregate level.

We assume that labor is supplied inelastically in each region r. Labor-market equilibrium in region r is defined by

$$\sum_{i\in\Omega_r} L_{si} = \bar{L}_{sr}, \qquad \sum_{i\in\Omega_r} L_{ui} = \bar{L}_{ur} \tag{4}$$

where $\bar{L}_{sr}$ and $\bar{L}_{ur}$ denote, respectively, the (exogenous) supply of skilled and unskilled labor in region r. Finally, the numéraire of the model is the final-good price ($P = 1$) and, at equilibrium, aggregate income is equal to the value of final output:

$$PY = \left(\frac{\varepsilon}{\varepsilon - 1}\right)\sum_r \left(w_{sr}\bar{L}_{sr} + w_{ur}\bar{L}_{ur}\right) \tag{5}$$

3.2 Firm Heterogeneity and Aggregate Elasticity of Substitution

To start with, we focus on the impact of a change in labor supply on the relative wage of skilled labor to unskilled labor in a region. Differentiating the conditional labor demand and marginal cost equations and using the equilibrium conditions on the regional labor markets,[8] we get

$$(\delta-\varepsilon)\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right)\sum_{i\in\Omega_r} s_{si}\lambda_{sri} + (\delta-\varepsilon)\,\mathrm{d}\ln w_{ur} - \delta\,\mathrm{d}\ln w_{sr} + \mathrm{d}\ln Y = \mathrm{d}\ln \bar{L}_{sr} \tag{6}$$

$$(\delta-\varepsilon)\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right)\sum_{i\in\Omega_r} s_{si}\lambda_{uri} + (\delta-\varepsilon)\,\mathrm{d}\ln w_{ur} - \delta\,\mathrm{d}\ln w_{ur} + \mathrm{d}\ln Y = \mathrm{d}\ln \bar{L}_{ur} \tag{7}$$

where s_{si} is the share of skilled employment in the total wage cost of firm i. $\lambda_{sri} = L_{si}/\bar{L}_{sr}$ denotes the share of firm i in region r's total skilled employment, and $\lambda_{uri} = L_{ui}/\bar{L}_{ur}$ is the equivalent definition for unskilled employment. Taking the difference between these two equations allows us to eliminate Y and leads to

$$\left[\sum_{i\in\Omega_r} s_{si}(\lambda_{sri}-\lambda_{uri})(\delta-\varepsilon) - \delta\right]\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) = \mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) \tag{8}$$

This allows us to state Proposition 1.

Proposition 1 *If resident workers are not mobile across regions and if there are small shocks to labor supply in all regions, the relative wage w_{sr}/w_{ur} in a region r depends only on the change in the relative labor supply in the same region, L_{sr}/L_{ur}. Moreover, this change is given by*

$$\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) = -\frac{1}{\sigma_r}\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right), \qquad \sigma_r = (1-\Psi_r)\delta + \Psi_r\varepsilon$$

where σ_r is the aggregate elasticity of substitution between skilled and unskilled labor at the regional

[8] The full set of differentiated equations of the model is given in the Appendix. All parameters of the differentiated model are summarized in Table 3.

level and Ψ_r *is defined as*

$$\Psi_r = \frac{\mathrm{Var}_r(s_{si})}{\theta_{sr}(1-\theta_{sr})},$$

where $\mathrm{Var}_r(s_{si}) = \sum_{i\in\Omega_r} s_{ir}(s_{si} - \theta_{sr})^2$ *and* θ_{sr} *is the mean skilled wage share in region* r*.*

From equation (8) and Proposition 1, it is clear that is suffices to show that $\sum_{i\in\Omega_r} s_{si}(\lambda_{sri} - \lambda_{uri}) = \frac{\mathrm{Var}_r(s_{si})}{\theta_{sr}(1-\theta_{sr})}$. See Appendix A.3.

This proposition establishes two main results. First, if there are small labor supply shocks in all regions, the relation between the relative skill supply and the relative wage in a region r does not depend on what happens in other regions. This result implies in particular that the elasticity of the relative wage with respect to relative labor supply is equal to the inverse of the aggregate elasticity of substitution between skilled and unskilled labor in the region. We define the aggregate (regional) elasticity of substitution between skilled and unskilled labor as[9]

$$\sigma_r = \frac{\mathrm{d}\ln(L_{sr}/L_{ur})}{\mathrm{d}\ln(w_{sr}/w_{ur})}, \text{ where } L_{sr} = \sum_{i\in\Omega_r} L_{si} \text{ and } L_{ur} = \sum_{i\in\Omega_r} L_{ui} \tag{9}$$

Second, the *aggregate* elasticity of substitution between skilled and unskilled labor at the regional level, σ_r, is a weighted average of the elasticity of substitution between skilled and unskilled labor *at the firm level*, δ, and the elasticity of substitution between the intermediate goods produced by firms, ε. The weight depends on the degree of heterogeneity between firms in region r, as measured by the (normalized) variance of the skilled share in firms' wage costs.

If there is no firm heterogeneity in the region ($\Psi_r = 0$), the aggregate elasticity σ_r is equal to the firm-level elasticity (δ). With increasing heterogeneity, the aggregate elasticity σ_r gets closer to the elasticity of substitution between intermediate inputs, ε. This result captures Rybczynski-type effects: an increase in the supply of skilled labor in region r increases the output of skill-intensive firms relative to other firms in the region. If the outputs of firms are highly substitutable (high ε), the labor supply shock can be absorbed mostly by a change in the output mix in the region and has little impact on relative wages. In this case, the adjustment takes place mostly between firms (as captured by ε) and only slightly within firms (as captured by δ).

[9] For the reader interested in the intricacies of the concepts of elasticity of substitution and complementarity, a few comments are in order. In Proposition 1, we are concerned with the impact of a change in relative labor supply on relative wages. This question is closely related to the elasticity of complementarity (Sato and Koizumi, 1973). In the case of two inputs, the elasticity of complementarity (which is defined as the elasticity of the technical rate of substitution between two inputs with respect to the relative quantity of these inputs) is the inverse of the elasticity of substitution. In the case of more than two inputs, there are a large number of definitions of partial elasticities of substitution and complementarity, and this result does not hold in general. However, in our model, the nested CES structure implies that the result of the two-factor case carries over to our setting as long as we consider two factors within a same nest (which is the case for skilled and unskilled labor within a region). To be more precise, the definition of the aggregate elasticity of substitution given in equation (9) is a two-factor two-price elasticity of substitution (TTES, see Chambers, 1988). This class of elasticities can be written as the weighted average of the Morishima elasticities of substitution between the two factors (Chambers, 1988, p.97). In our model, the Morishima elasticities are symmetric (since $\mathrm{d}\ln(L_{sr}/L_{ur})/\mathrm{d}\ln w_{sr} = \mathrm{d}\ln(L_{ur}/L_{sr})/\mathrm{d}\ln w_{ur}$), implying that the TTES are identical to the Morishima elasticity. As Blackorby and Russell (1989) argue convincingly, the Morishima elasticity is indeed the relevant concept in the multifactor case.

Table 3: Parameters of the Differentiated Model

s_i	$=$	$(p_i y_i)/(PY)$	share of firm i in the value of aggregate output
s_{si}	$=$	$(w_{sr} L_{si})/(c_i y_i)$	share of skilled employment in total wage cost of firm i
s_{ir}	$=$	$(c_i y_i)/(\sum_{l \in \Omega_r} c_l y_l)$	share of firm i in wage costs of region r
s_r	$=$	$\sum_{i \in \Omega_r} s_i$	share of region r in aggregate output
λ_{sri}	$=$	$L_{si}/\bar{L}_{sr}$	share of firm i in skilled employment in region r
λ_{uri}	$=$	$L_{ui}/\bar{L}_{ur}$	share of firm i in unskilled employment in region r
λ_{ur}	$=$	$\bar{L}_{ur}/(\bar{L}_{sr} + \bar{L}_{ur})$	share of unskilled employment in total employment of region r
θ_{sr}	$=$	$\sum_{i \in \Omega_r} s_{si} s_{ir}$	average share of skilled employment in wage cost of region r
θ_s	$=$	$\sum_r s_r \theta_{sr}$	average share of skilled employment in aggregate wage cost
θ_{ms}	$=$	$\sum_r \alpha_{sr} \theta_{msr}$	average share of migrants in aggregate skilled wage cost
θ^*_{ms}	$=$	$\sum_r \alpha_{sr} \theta^*_{msr}$	average share of "new" migrants in aggregate skilled wage cost
θ_{msr}	$=$	$w_{sr} M_{sr}/(w_{sr} \bar{L}_{sr})$	share of migrants in skilled wage cost of region r
θ^*_{msr}	$=$	$w_{sr} M^*_{sr}/(w_{sr} \bar{L}_{sr})$	share of "new" migrants in skilled wage cost of region r
α_{sr}	$=$	$s_r \theta_{sr}/\theta_s$	share of region r in total skilled wage cost

Notes: This table shows the relevant parameters of the differentiated model. Note that $c_i y_i = w_{sr} L_{si} + w_{ur} L_{ui}$ and $\sum_{i \in \Omega_r} c_i y_i = w_{sr} \bar{L}_{sr} + w_{ur} \bar{L}_{ur}$

To see why the variance of skill shares in Proposition 1 is normalized by the factor $\theta_{sr}(1 - \theta_{sr})$, assume at the extreme that firms use either skilled or unskilled labor (but never both). In this case, the substitution between skilled and unskilled labor only happens between firms and we should have $\sigma_r = \varepsilon$. It is indeed straightforward to see that in this case $\text{Var}_r(s_{si}) = \theta_{sr}(1 - \theta_{sr})$ and therefore $\Psi_r = 1$.[10]

3.3 Changes in Regional Labor Supply and Wage Rates

Proposition 1 focuses on the reaction of *relative* wages to a change in labor supply. We now turn to the impact of a change in labor supply on the *level* of real wages in a region. We start again with equations (6) and (7) in order to calculate the impact of changes in labor supply on the output of each region r. Using the results from equations (A.5) and (A.6) in the appendix, we obtain

$$\mathrm{d}\ln \bar{L}_{sr} = (\delta - \varepsilon)\, \mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) \frac{\sum_{i \in \Omega_r} s_{ir} s_{si}^2}{\theta_{sr}} + (\delta - \varepsilon)\, \mathrm{d}\ln w_{ur} - \delta\, \mathrm{d}\ln w_{sr} + \mathrm{d}\ln Y \tag{10}$$

$$\mathrm{d}\ln \bar{L}_{ur} = (\delta - \varepsilon)\, \mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) \frac{\theta_{sr} - \sum_{i \in \Omega_r} s_{ir} s_{si}^2}{1 - \theta_{sr}} + (\delta - \varepsilon)\, \mathrm{d}\ln w_{ur} - \delta\, \mathrm{d}\ln w_{ur} + \mathrm{d}\ln Y \tag{11}$$

where s_{ir} is the share of firm i in the value of output in region r. Now consider the weighted average of the change in skilled and unskilled labor supply in region r, and denote this change by $\mathrm{d}\ln Y_r$

[10] To see this, assume that a share θ_{sr} of firms uses only skilled labor (and therefore $s_{si} = 1$) and a share $(1 - \theta_{sr})$ of firms only uses unskilled labor (and therefore $s_{si} = 0$). Therefore, the average skilled-wage share in the region is θ_{sr}, and the variance is given by $\text{Var}_r(s_{si}) = \theta_{sr}(1 - \theta_{sr})^2 + (1 - \theta_{sr})(0 - \theta_{sr})^2 = \theta_{sr}(1 - \theta_{sr})$, which is the maximum value that the variance can attain.

(the interpretation of Y_r will become clearer below)

$$\mathrm{d}\ln Y_r = \theta_{sr}\,\mathrm{d}\ln \bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln \bar{L}_{ur} \tag{12}$$

Using equations (10) and (11), the relative change in Y_r can be written as

$$\mathrm{d}\ln Y_r = -\varepsilon\,[\theta_{sr}\,\mathrm{d}\ln w_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln w_{ur}] + \mathrm{d}\ln Y \tag{13}$$

Similarly to $\mathrm{d}\ln Y_r$, we can define the change in the price of output in region r as the weighted average of the change in regional wages of skilled and unskilled labor

$$\mathrm{d}\ln P_r = \theta_{sr}\,\mathrm{d}\ln w_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln w_{ur} \tag{14}$$

Equation (13) can now be further simplified to

$$\mathrm{d}\ln Y_r = -\varepsilon\,\mathrm{d}\ln P_r + \mathrm{d}\ln Y \tag{15}$$

Considering that the change in relative wages is known from Proposition 1, we are now in a position to solve these equations for the level of the skilled wage. Note that from (14), we can express $\mathrm{d}\ln P_r$ as a function of the skilled wage and the skilled-to-unskilled wage ratio: $\mathrm{d}\ln P_r = \mathrm{d}\ln w_{sr} - (1-\theta_{sr})\,\mathrm{d}\ln(w_{sr}/w_{ur})$. On the other hand, we can solve equation (15) for P_r and combine the two equations to yield

$$\mathrm{d}\ln w_{sr} = \frac{1}{\varepsilon}(\mathrm{d}\ln Y - \mathrm{d}\ln Y_r) + (1-\theta_{sr})\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) \tag{16}$$

Using Proposition 1 to eliminate the relative wage in equation (16), rearranging equation (12) to obtain $(1-\theta_{sr})\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) = \mathrm{d}\ln \bar{L}_{sr} - \mathrm{d}\ln Y_r$, we finally get

$$\mathrm{d}\ln w_{sr} = \frac{1}{\varepsilon}\,\mathrm{d}\ln Y + \left(\frac{1}{\sigma_r} - \frac{1}{\varepsilon}\right)\mathrm{d}\ln Y_r - \frac{1}{\sigma_r}\,\mathrm{d}\ln \bar{L}_{sr} \tag{17}$$

where σ_r is given by Proposition 1, $\mathrm{d}\ln Y_r$ is defined by equation (12), and $\mathrm{d}\ln Y$ can be obtained by differentiating equation (5) (see the last equation in Appendix A.2). To solve for the unskilled wage rate, we can rearrange (14) to express $\mathrm{d}\ln P_r$ as a function of the unskilled wage and the relative wage ratio, $\mathrm{d}\ln P_r = \mathrm{d}\ln w_{ur} + \theta_{sr})\,\mathrm{d}\ln(w_{sr}/w_{ur})$, and follow the same steps as for the skilled wage rate. These results are summarized in Proposition 2.

Proposition 2 *If resident workers are not mobile across regions and if there are small shocks to labor supply in all regions, the changes in the skilled and unskilled wage rates are given by*

$$\mathrm{d}\ln w_{sr} = \frac{1}{\varepsilon}(\mathrm{d}\ln Y - \mathrm{d}\ln Y_r) + \frac{1}{\sigma_r}\left(\mathrm{d}\ln Y_r - \mathrm{d}\ln \bar{L}_{sr}\right) \tag{18}$$

$$\mathrm{d}\ln w_{ur} = \frac{1}{\varepsilon}(\mathrm{d}\ln Y - \mathrm{d}\ln Y_r) + \frac{1}{\sigma_r}\left(\mathrm{d}\ln Y_r - \mathrm{d}\ln \bar{L}_{ur}\right) \tag{19}$$

where

$$\begin{aligned}
\mathrm{d}\ln Y &= \sum_r s_r \left[\theta_{sr}\,\mathrm{d}\ln \bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln \bar{L}_{ur}\right], \\
\mathrm{d}\ln Y_r &= \theta_{sr}\,\mathrm{d}\ln \bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln \bar{L}_{ur},
\end{aligned}$$

and σ_r is defined in Proposition 1.

Proposition 2 shows that there are two effects at work in the determination of regional real wage levels. First, the *within-region* substitution effect (as described by Proposition 1) obviously plays a role in determining real wage levels. It is captured by the last term in equations (18) and (19). If the labor supply shock in region r increases skilled labor more than unskilled labor, the relative increase in Y_r is smaller than the increase in $\bar{L}_{sr}$ (and greater than the increase in $\bar{L}_{ur}$) and the substitution effect tends to decrease the skilled real wage in region r (and increase the unskilled real wage). The magnitude of this effect depends on the inverse of the aggregate regional elasticity of substitution, σ_r, as defined in Proposition 1.

Second, there is the *between-region* effect, as described by the first term on the righthand side of equations (18) and (19). If the labor supply shock is greater in other regions than in region r, the relative increase in Y outweighs the increase in Y_r; this tends to increase real wages in region r, both for skilled and unskilled workers. The magnitude of this effect depends on the inverse of the elasticity of substitution, ε, between firms (and thus between products of different regions). In this case there is a *complementarity effect* between regions: the increase in labor supply in other regions raises overall demand in the country and therefore also in region r, raising wages in region r.

Proposition 2 has another implication that is useful for the economic interpretation of the model in the context of regional immigration. When we define $\mathrm{d}\ln Y_r$ in equation (12) above, this amounts to define Y_r as the aggregate output in region r. In this context, the aggregate production function of the model can be reinterpreted as depicted in Panel A of Figure 3. At the lower level, firms combine skilled and unskilled labor using a CES technology with an elasticity of substitution δ (and heterogeneous relative skill productivity). At the next level, the output of firms in a same region r is combined using a CES with elasticity ε.[11] At the upper level, aggregate output at the national level is produced by combining the regional outputs, using a CES with elasticity ε.

Proposition 2 tells us that the two lower levels can be combined into one level, as depicted in Panel B of Figure 3. In this (equivalent) representation of the model, there is a representative firm in each region r which uses all skilled and unskilled labor in the region and is characterized by a CES technology with elasticity of substitution σ_r. Note that this "reduced-form" representation of

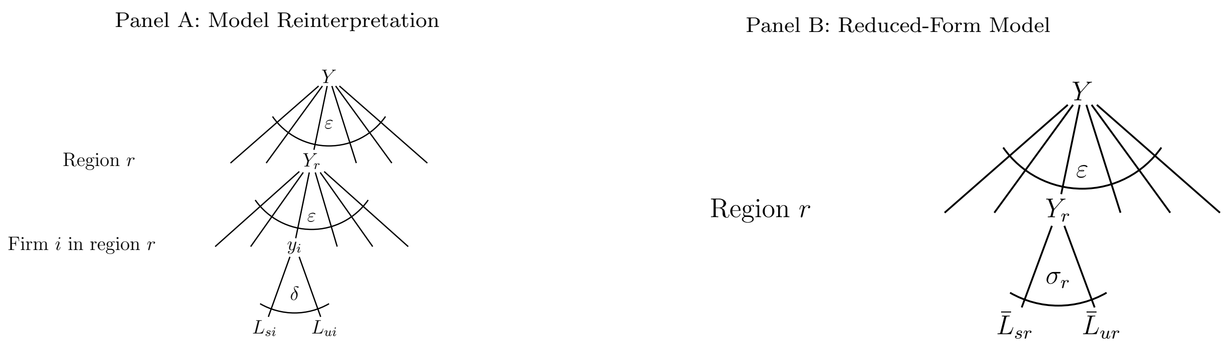
[11] More formally, the variable Y_r can be given a precise interpretation by rewriting the final-good production function (1) as a nested CES function:

$$Y = \left(\sum_r \tilde{\beta}_r Y_r^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \qquad Y_r = \left(\sum_{i\in\Omega_r} \bar{\beta}_i y_i^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}$$

where $\bar{\beta}_i$ and $\tilde{\beta}_r$ are related to the original parameters β_i as follows: $\bar{\beta}_i\tilde{\beta}_r = \beta_i$. The dual price equation (A.2) can be similarly decomposed.

the model is much easier to handle than the original form, because it is based on a *separable* nested CES production function (i.e., regional labor supplies $\bar{L}_{sr}$ and $\bar{L}_{ur}$ enter the CES tree at a single location). This version of the model greatly facilitates the analysis of regional immigration shocks below.

Figure 3: Alternative Structures of the Aggregate Production Function



Note: This figure provides alternative depictions of the aggregate production function that are equivalent to the main one.

3.4 The Impact of Immigration on Average Native Wages

In this section, we study how immigration affects natives' wages, focusing on the impact of skilled immigration on the average wage of skilled natives at the national level.[12] The latter is a weighted average of regional skilled wages and can differ from the average wage of skilled migrants if skilled natives and migrants are not identically distributed across regions.

To distinguish migrants and natives among skilled and unskilled workers, we adopt the following notations. Total supply of skilled labor in region r is given by $\bar{L}_{sr} = N_{sr} + M_{sr}$, where M_{sr} (N_{sr}) denotes the exogenous supply of skilled migrants (natives) in region r. The total number of skilled migrants in the country is $M_s = \sum_r M_{sr}$.[13] We consider a small increase in the number of skilled migrants in the country and assume that these "new" skilled migrants are distributed among regions r in the proportions m^*_{sr}. The following notation will turn out to be useful. Denote by M^*_{sr} the number of "old" migrants who would live in region r if all "old" migrants were distributed among regions according to the proportions m^*_{sr} (i.e., $M^*_{sr} = m^*_{sr} M_s$). With these definitions, the relative variation of skilled labor supply in region r can now be connected to the relative increase in skilled immigration (where the change in the number of skilled immigrants, $\mathrm{d}M_s$, is expressed relative to the total stock of skilled labor, L_s):

$$\mathrm{d}L_{sr} = m^*_{sr}\,\mathrm{d}M_s \quad \Rightarrow \quad \frac{\mathrm{d}L_{sr}}{L_{sr}} = \frac{M^*_{sr}}{M_s}\frac{L_s}{L_{sr}}\frac{\mathrm{d}M_s}{L_s} = \frac{(M^*_{sr}/L_{sr})}{(M_s/L_s)}\frac{\mathrm{d}M_s}{L_s} = \frac{\theta^*_{msr}}{\theta_{ms}}\frac{\mathrm{d}M_s}{L_s} \tag{20}$$

[12] As stated before, we assume that resident workers are not mobile between regions. We discuss the case of perfect mobility in the next subsection.

[13] Similarly, supply of unskilled labor in region r is given by $\bar{L}_{ur} = N_{ur} + M_{ur}$, where N_{sr} (M_{sr}) denotes the exogenous supply of unskilled natives (migrants) and $M_s = \sum_r M_{sr}$ is the total number of unskilled migrants in the country.

where θ^*_{msr} denotes the share of "new" migrants in the skilled wage cost of region r[14] and θ_{ms} is the share of migrants in aggregate skilled wages (all parameters are summarized in Table 3).

We now turn to analyzing the impact of immigration on the average wage of skilled natives, w_{ns}. Because skilled natives and migrants earn the same wage within the same region, we rely on a national wage of natives that is constructed as a weighted average of skilled regional wage rates, w_{sr}. As shown in Appendix A.4, the relative change in w_{ns} is given by

$$\mathrm{d}\ln w_{ns} = \sum_r \frac{\alpha_{sr}(1-\theta_{msr})}{1-\theta_{ms}}\,\mathrm{d}\ln w_{sr} \tag{21}$$

where θ_{msr} is the share of "old" migrants in the skilled wage cost of region r and α_{sr} denotes the share of region r in the total skilled wage cost. The potential role of the network effect is already apparent in the last two equations. A strong network effect implies that the distribution of "new" migrants is close to the distribution of "old" migrants, which would be reflected in a strong correlation between θ^*_{msr} and θ_{msr}.

We are now ready to establish the link between skilled immigration and the average wage of skilled natives. Equation (18) from Proposition 2 provides the missing link between equations (20) and (21), leading to Proposition 3.

Proposition 3 *If there is a small increase in skilled immigration and if there is no mobility of workers between regions, the relative change in the average wage of skilled native workers is given by*

$$\begin{aligned}\mathrm{d}\ln w_{ns} \quad = \quad & \left(\frac{1}{\varepsilon}\sum_r \frac{\alpha_{sr}[\theta_s(1-\theta_{ms})\theta^*_{ms} - \theta_{sr}(1-\theta_{msr})\theta^*_{msr}]}{\theta_{ms}(1-\theta_{ms})}\right. \\ & \left. - \sum_r \frac{1}{\sigma_r}\frac{\alpha_{sr}(1-\theta_{sr})(1-\theta_{msr})\theta^*_{msr}}{\theta_{ms}(1-\theta_{ms})}\right)\frac{\mathrm{d}M_s}{L_s},\end{aligned} \tag{22}$$

θ^*_{ms} is the share of new migrants in aggregate skilled wages.

See Appendix A.4.

An economic interpretation of Proposition 3 is not straightforward, since it involves regional heterogeneity in three dimensions: skill intensity (θ_{sr}), distribution of old migrants (θ_{msr}), and distribution of new migrants (θ^*_{msr}). To facilitate the interpretation, we present two corollaries that analyze instructive cases. First, we focus on the skill heterogeneity across regions, assuming that the migrant-native composition of skilled labor is identical in all regions and that the increase in skilled migrants is evenly spread across regions. Second, we focus on the native-migrant composition heterogeneity across regions, assuming that the skill composition is the same in all regions, and we

[14] More precisely, θ^*_{msr} denotes the share that "new" migrants would represent in the skilled wage cost of region r if the stock of "old" migrants were distributed across regions according to the proportions of new migrants. This definition of the cost share of "new" skilled migrants might seem cumbersome, but it ensures that it has a comparable scale to the cost share of "old" skilled migrants, θ_{msr}. This facilitates the interpretation of the covariance between the two parameters.

analyze an increase in skilled migrants whose regional distribution is related to the initial settlement of old migrants. In both corollaries, we assume that regions have the same firm heterogeneity ($\Psi_r = \Psi$ and therefore $\sigma_r = \sigma$).

Corollary 3.1 (Skill composition heterogeneity across regions) *Assume $\theta_{msr} = \theta_{ms}$, $\theta^*_{msr} = \theta^*_{ms}$, and $\sigma_r = \sigma$. If there is no mobility across regions, a small increase in skilled immigration that is evenly spread across regions leads to the following relative change in the average wage of skilled native workers:*

$$\mathrm{d}\ln w_{ns} \quad = \quad -(1-\theta_s)\left(\frac{\Upsilon_s}{\varepsilon} + \frac{1-\Upsilon_s}{\sigma}\right)\frac{\theta^*_{ms}}{\theta_{ms}} \cdot \frac{\mathrm{d}M_s}{L_s} \tag{23}$$

where

$$\Upsilon_s = \frac{\mathrm{Var}(\theta_{sr})}{\theta_s(1-\theta_s)} \quad \textit{and} \quad \mathrm{Var}(\theta_{sr}) = \sum_r s_r(\theta_{sr}-\theta_s)^2.$$

See Appendix A.4.

According to the assumptions of Corollary 3.1, the arrival of new skilled migrants increases the skilled labor force in the same proportion in all regions. Corollary 3.1 shows that the impact of this labor supply shock on the average wage of skilled natives is a combination of a within-region substitution effect between skilled and unskilled labor, and a change in the output mix across regions. The relative importance of these two mechanisms depends crucially on the heterogeneity of regions in terms of skill intensity (as described by Υ_s, the normalized variance of the mean skilled wage share in region r). At one extreme, if all regions have the same skill intensity ($\theta_{sr} = \theta_s$ and $\Upsilon_s = 0$), only the substitution effect operates and the average skilled wage of natives decreases by a factor $\left(\frac{1-\theta_s}{\sigma}\right)$. Note that this effect is identical to the aggregate impact of skilled immigration in the case of perfect mobility of labor between regions (see Proposition 2' below).[15] At the other extreme, there is complete spatial segregation between skilled and unskilled workers, which implies maximum skill heterogeneity of regions ($\Upsilon_s = 1$). In this case, skilled immigration increases the output of firms located in skill-intensive regions and the impact of wages depends only on the elasticity of substitution, ε, between firms' outputs.

In the more general, intermediate case, the impact of immigration on wages is equal to a weighted average of the two effects, where Υ_s represents the relative weight. The mechanism that drives this regional skill heterogeneity effect of Corollary 3.1 is related to (but distinct from) the Rybczynski-type effects that take place *within* regions (see our discussion of Proposition 1). The difference resides in the fact that labor is perfectly *mobile* between firms *within a region* (Rybczynski-type effects in Proposition 1) but *immobile between regions* (regional skill heterogeneity effect of Corollary 3.1).

We now turn to the second corollary, which focuses on the heterogeneity across regions in terms of native-migrant composition, assuming that the skill composition is the same in all regions.

Corollary 3.2 (Native-migrant composition heterogeneity across regions.) *Assume $\theta_{sr} = \theta_s$ and $\sigma_r = \sigma$. If there is no mobility across regions, a small increase in skilled immigration that is*

[15] With a suitable choice of units (e.g., initial wage rates normalized to one), the ratio $(\theta^*_{ms}/\theta_{ms})$ is equal to 1.

spread across regions proportional to θ_{ms} leads to the following relative change in the average wage of skilled native workers:

$$\mathrm{d}\ln w_{ns} = \left[-\frac{(1-\theta_s)}{\sigma}\cdot\frac{\theta^*_{ms}}{\theta_{ms}} + \Lambda_{ms}\left(\frac{\theta_s}{\varepsilon}+\frac{(1-\theta_s)}{\sigma}\right)\right]\frac{\mathrm{d}M_s}{L_s} \tag{24}$$

where the normalized covariance between θ_{msr} and θ^*_{msr} is defined as $\Lambda_{ms} = \frac{\mathrm{Cov}(\theta_{msr},\theta^*_{msr})}{\theta_{ms}(1-\theta_{ms})}$, and $\mathrm{Cov}(\theta_{msr},\theta^*_{msr}) = \sum_r \alpha_{sr}(\theta_{msr}-\theta_{ms})(\theta^*_{msr}-\theta^*_{ms})$.

See Appendix A.4.

Also in this case, immigration has two distinct effects under the assumptions of Corollary 3.2. First, skilled immigration decreases the average wage of skilled natives by a factor $\frac{1-\theta_s}{\sigma}$. Second, the spatial clustering of immigrants tends to attenuate the negative wage effect of immigration for natives, represented by the second element in square brackets of Equation (24). To see why, consider again two extreme scenarios. If old skilled migrants are equally distributed across regions, then $\mathrm{Var}(\theta_{msr}) = 0$, which implies $\Lambda_{ms} = 0$,[16] and the effect of skilled immigration is driven only by the substitution between skilled and unskilled labor within firms, as captured by the term $\frac{1-\theta_s}{\sigma}$. Instead, if there is full spatial segregation between skilled natives and old migrants and new migrants settle in the same regions as old migrants, then $\mathrm{Var}(\theta_{msr})$ and $\mathrm{Var}(\theta^*_{msr})$ are close to their maximal value $\theta_{ms}(1-\theta_{ms})$ and Λ_{ms} is close to 1. This means that reallocations of output across regions take place and reduce the negative effect of immigration. The reason is that skilled wages decrease only in regions to which new skilled migrants are allocated, thus sheltering skilled native wages. In this extreme case the reallocation effect is so strong as to offset the negative effect given by $\frac{1-\theta_s}{\sigma}$. More generally, the size of this effect depends on the migrant/native segregation across regions, i.e., $Var(\theta_{msr})$ and on the location choice of new migrants, i.e., the size of $Cov(\theta_{msr},\theta^*_{msr})$.

Overall, the message that arises from Corollaries 3.1 and 3.2 is that the greater the heterogeneity in the distribution of skills (Υ_s) and migrants across regions (and the more new migrants settle where old migrants are, as reflected by a higher Λ_{ms}), the stronger the reallocations across regions and the smaller the negative effect of immigration on natives' wages. Moreover, the mechanism described by Corollary 3.2 explains why (un)skilled natives and (un)skilled migrants might appear as imperfect substitutes in the aggregate, though they are perfect substitutes at the firm level.

3.5 The Case of Perfect Mobility

In this subsection, we analyze the role of labor mobility in the context of our model. This is a crucial issue in understanding the impact of immigrants on natives' wages. As highlighted earlier in our stylized facts, we find that in the Swiss context immigration does not affect residents' location choice. Therefore, an immigration shock does not induce a reallocation of workers across regions, thus supporting the hypothesis of segmented labor markets. However, it is interesting to analyze how the model would change in the case of perfect mobility. Assuming that workers can freely move

[16] This is because $\mathrm{Cov}(\theta_{msr},\theta^*_{msr}) = \mathrm{Corr}(\theta_{msr},\theta^*_{msr})\sqrt{\mathrm{Var}(\theta_{msr})\mathrm{Var}(\theta^*_{msr})}$.

across regions, we have that the supply of skilled labor, $\bar{L}_s$, and unskilled labor, $\bar{L}_u$ are exogenous at the national level and labor market equilibrium conditions (4) are replaced by

$$\sum_{i\in\Omega} L_{si} = \bar{L}_s, \qquad \sum_{i\in\Omega} L_{ui} = \bar{L}_u \tag{25}$$

In this version of the model, regional wage rates are equalized across regions:

$$w_{sr} = w_s, \qquad w_{ur} = w_u \tag{26}$$

Proposition 1' describes the impact of a change in relative labor supply (at the national level) on relative wages. The aggregate elasticity of substitution is now defined at the national level, and the weight Ψ depends on the variance of firms' skill share in the entire country.

Proposition 1' (Perfect labor mobility between regions) *If labor is perfectly mobile between regions and if there are small shocks to labor supply at the national level, the relative wage w_s/w_u in the country depends on the change in the relative labor supply in the country, L_s/L_u, as follows:*

$$\mathrm{d}\ln\left(\frac{w_s}{w_u}\right) = -\frac{1}{\sigma}\,\mathrm{d}\ln\left(\frac{\bar{L}_s}{\bar{L}_u}\right), \qquad \sigma = (1-\Psi)\delta + \Psi\varepsilon$$

where σ is the aggregate elasticity of substitution between skilled and unskilled labor at the national level and Ψ is defined as

$$\Psi = \frac{\mathrm{Var}(s_{si})}{\theta_s(1-\theta_s)}$$

where $\mathrm{Var}(s_{si}) = \sum_{i\in\Omega} s_i(s_{si}-\theta_s)^2$, s_i denotes the share of firm i in the value of aggregate output, and θ_s is the mean skilled wage share in the country.

Proposition 1' is a straightforward application of Proposition 1 to a single region.

Switching to wage levels, we can consider the following corollary.

Proposition 2' (Perfect labor mobility between regions) *If labor is perfectly mobile between regions and if there are small shocks to labor supply at the national level, the changes in the skilled and unskilled wage rates at the national level are given by*

$$\mathrm{d}\ln w_s = -\left(\frac{1-\theta_s}{\sigma}\right)\mathrm{d}\ln\left(\frac{\bar{L}_s}{\bar{L}_u}\right) \tag{27}$$

$$\mathrm{d}\ln w_u = \left(\frac{\theta_s}{\sigma}\right)\mathrm{d}\ln\left(\frac{\bar{L}_s}{\bar{L}_u}\right) \tag{28}$$

where σ and θ_s are defined in Proposition 1'.

In a nutshell, if there is perfect mobility, the attenuating effect of the across-region reallocation of output is canceled and wages are set at the national level. Only the across-firm reallocation effect remains in place, meaning that the negative effect of immigration is attenuated only by expanding firms.

3.6 Extended Model: Human-Capital Externalities

We now examine how the benchmark model would change by assuming human-capital externalities at the regional level, in the spirit of Moretti (2004). This allows us to account for the widely documented productivity-enhancing effects of skilled immigration (e.g., Kerr and Lincoln, 2010; Ghosh et al., 2014; Hornung, 2014; Mayda et al., 2020; Mitaritonna et al., 2017; Beerli et al., 2021) and to reconcile our theoretical model with the empirical evidence for Switzerland presented in Beerli et al. (2021).

A crucial assumption in our approach is that firms are heterogeneous with respect to the importance of this externality. This assumption is founded on the observation that high-tech manufacturing firms and knowledge-intensive firms can be assumed to benefit more from a high density of skilled workers in the region.[17] If the human-capital externality has a stronger effect on the productivity of skill-intensive firms, we show below that the impact of skilled immigration on relative wages is attenuated and can even be reversed: skilled immigration can lead to an *increase* in the relative wages of skilled workers.

The economic mechanism underlying this result is as follows. Though we assume that the externality takes a skill-neutral form at the firm level, we show that the externality is skill-biased at the aggregate (regional) level if the externality operates predominantly in skill-intensive firms, i.e., if there is a positive covariance between the impact of the externality on a firm's productivity and its skill intensity. The aggregate skill bias of the externality can be sufficiently strong to imply that skilled immigration leads to an increase in the relative wage of skilled labor compared to unskilled labor in a region.

In this extended version of our model, we assume that total factor productivity A_i of a firm becomes endogenous in equation (2) and depends positively on the ratio of skilled to unskilled labor in the region where the firm operates:[18]

$$A_i = A_i^0 \left(\frac{\bar{L}_{sr}}{\bar{L}_{sr} + \bar{L}_{ur}} \right)^{\eta_i} \tag{29}$$

Two properties of this specification are noteworthy. First, the impact of the externality on productivity is not factor-biased, which implies that it does not affect the relative demand for skilled versus unskilled labor at the firm level. Second, the elasticity η_i is differentiated by firm and may be correlated with the skill composition of the firm's labor force. A positive correlation indicates that skill-intensive firms tend to be more sensitive to the regional externality.

We can now state an extended version of Proposition 1, taking regional externalities into account.

Proposition 1" *(Human-capital externalities). If there are regional externalities as defined in equation (29), if resident workers are not mobile across regions and if there are small shocks to*

[17] Beerli et al. (2021) provide evidence of positive effects of high-skill immigration in Switzerland on labor productivity in high-tech manufacturers and knowledge-intensive business service firms (Table 6, Panel A).

[18] As Moretti (2004a) notes, different microfoundations have been proposed in the literature that can generate this kind of externality (e.g., Lucas 1988, Acemoglu 1996).

labor supply in all regions, the relative wage w_{sr}/w_{ur} in a region r depends only on the change in the relative labor supply in the same region, L_{sr}/L_{ur}. Moreover, this change is given by

$$\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) = -\left(\frac{1-(\varepsilon-1)\lambda_{ur}\Phi_r}{\sigma_r}\right)\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right),$$

where σ_r is the aggregate elasticity of substitution between skilled and unskilled labor at the regional level (defined in Proposition 1), $\lambda_{ur} = \bar{L}_{ur}/(\bar{L}_{sr} + \bar{L}_{ur})$, and Φ_r is given by

$$\Phi_r = \frac{\mathrm{Cov}_r(s_{si}, \eta_i)}{\theta_{sr}(1-\theta_{sr})} = \sum_{i\in\Omega_r} s_{ir}\frac{(s_{si}-\theta_{sr})(\eta_i-\bar{\eta}_r)}{\theta_{sr}(1-\theta_{sr})}$$

where $\bar{\eta}_r = \sum_{i\in\Omega_r} s_{ir}\eta_i$.

Proof: see Appendix A.6.

Regarding Proposition 1”, two remarks are in order. First, it is possible that the relative wage of skilled workers rises if the covariance between a firm’s skill intensity s_{si} and its elasticity η_i with respect to the externality is positive and sufficiently large. More precisely, there is a positive relationship between relative labor supply and relative wages in region r if $(\varepsilon - 1)\Phi_r > 1$. Second, if the elasticity of the externality is not correlated with firms’ skill intensity ($\Phi_r = 0$), Proposition 1” reduces to Proposition 1.

The economic intuition for Proposition 1” can be explained by considering the case of skilled immigration in a given region. If the positive externality produced by this type of immigration affects mainly skill-intensive firms (positive covariance Φ_r between s_{si} and η_i), the marginal costs of the latter will decrease and demand will shift toward intermediate goods produced by these firms. Therefore, demand for skilled labor tends to increase in the aggregate, driving skilled wages up relative to unskilled wages. If intermediate goods are close substitutes (high ε in combination with a positive Φ_r), this reallocation effect can be so strong that relative skilled wages increase with skilled immigration.

Further intuition can be gained by considering an aggregate representation of the model with regional externalities, along the lines of the “reduced form” of the model represented in Panel B of Figure 3. With regional externalities, the aggregate (regional) production function described by the lower level in Panel B of Figure 3 can be written as

$$Y_r = A_r\left[b_{sr}\left(B_r\bar{L}_{sr}\right)^{\frac{\omega_r-1}{\omega_r}} + b_{ur}\bar{L}_{ur}^{\frac{\omega_r-1}{\omega_r}}\right]^{\frac{\omega_r}{\omega_r-1}} \tag{30}$$

In this representation, the regional externality has two components: a skill-neutral component, A_r, and a skill-biased component, B_r.

$$A_r = A_r^0\left(\frac{\bar{L}_{sr}}{\bar{L}_{sr}+\bar{L}_{ur}}\right)^{\kappa_r}, \qquad B_r = B_r^0\left(\frac{\bar{L}_{sr}}{\bar{L}_{sr}+\bar{L}_{ur}}\right)^{\gamma_r}$$

How can the reduced-form parameters ω_r, κ_r, and γ_r be related to the structural parameters of the extended model? We show in Appendix A.7 that the aggregate representation of the production process at the regional level yields the same impact of immigration on regional output, prices, and (relative) wages as the structural model with regional externalities if

$$\omega_r = \sigma_r, \qquad \gamma_r = \Phi_r \left(\frac{\varepsilon - 1}{\sigma_r - 1} \right) \qquad \text{and} \qquad \kappa_r = \bar{\eta}_r - \theta_{sr} \Phi_r \left(\frac{\varepsilon - 1}{\sigma_r - 1} \right)$$

First, the aggregate elasticity of substitution between skilled and unskilled labor is given by σ_r, as in the base version of the model depicted in Panel B of Figure 3. Second, the skill-biased component of the externality depends on the covariance Φ_r between a firm's skill intensity s_{si} and its sensitivity to the regional externality, η_i. If these two parameters are positively correlated, the externality is skill-biased at the aggregate level (i.e., $\gamma_r > 0$), though the externality is skill-neutral at the firm level. The economic intuition for this result is related to the reallocation effects between firms described above (for the case of a positive covariance Φ_r). In this case, skilled immigration leads to a reallocation of skilled labor within a region towards skill-intensive firms, which benefit more from the externality (due to higher η_i). At the aggregate level, this reallocation shows up as a skill-biased effect of the externality. Third, the externality also has a skill-neutral effect at the aggregate level, which is represented by κ_r. The combination of the skill-biased (γ_r) and skill-neutral (κ_r) components determines the overall effect of the externality. For example, the externality is skill-neutral at the aggregate level ($\gamma_r = 0$ and $\kappa_r = \bar{\eta}_r$) if firms' skill intensity and sensitivity to the externality are uncorrelated ($\Phi_r = 0$).[19]

One final remark regarding this extended model: we consider only the case where workers are not mobile across regions. Perfect mobility of workers across regions may not be compatible with a unique equilibrium if the (skill-bias of the) externality is sufficiently strong.

4 Quantification Exercises

In this section, we quantify the main mechanisms of the benchmark model and of the augmented model with human-capital externalities. In both cases, we perform two counterfactual exercises based on an exogenous increase in labor supply due to immigration.

In the first, which we denote as "Diaspora Shock," we analyze an increase in labor supply due to immigration that is driven by a diaspora effect. In other words, we allocate the total inflow of immigrants into Switzerland of a certain nationality observed in the 2004–2010 period based on the share of migrants of the same nationality in each Swiss MS-region in 1980. This allocation rule exploits the fact that migrants tend to settle in locations in which immigrants of the same nationality already reside (e.g., Card, 2001). Panel A of Figure 4 depicts the labor supply increase across regions. As is evident, the range of the shock is quite important, with more than 40 MS-regions experiencing increases that are beyond 10% in terms of skilled labor and a similar number of regions increasing the number of unskilled workers by more than 2%. The skilled and unskilled

[19]Note that γ_r can be negative (if the covariance Φ_r is negative), which results in an "unskill-biased" effect of the externality at the aggregate level.

shocks do not overlap, and the relative labor supply increase is quite heterogeneous across regions, depending on the different settlement of skilled and unskilled migrant workers.

In the second counterfactual, which we denote as "AFMP Shock," we exploit differences in the timing and intensity of the implementation of the AFMP across regions in Switzerland described in Losa et al. (2012), Beerli et al. (2021) and Ariu (2022) to exogenously allocate foreign workers across MS-regions. Specifically, we exploit two sources of variation: (1) the AFMP lifted the restrictions against hiring cross-border workers (CBW) in the border regions (BR) before the central region (CR), i.e., 2004 versus 2007; (2) the intensity of the labor supply shock depended on the time distance from the border, because most CBW have a limited propensity to travel for long times to get to work (Beerli et al., 2021).

This setting allows for a simple difference-in-difference regression in which we compare the growth in the number of CBW across highly treated MS-regions (i.e., within 15 minutes from the border) and mildly treated ones (i.e., between 15 minutes and 30 minutes from the border) with respect to untreated ones (i.e., beyond 30 minutes from the border) before and after the AFMP was implemented. With this strategy, we can predict CBW inflows by skill type depending on the distance from the border and the AFMP timing. Panel B of Figure 4 shows that this shock is milder than the Diaspora Shock, and it substantially affects only regions close to the border, both for skilled and for unskilled workers. However, there are important differences across regions in the relative increase of skilled workers, meaning that even for regions close to the border, the capacity of attracting skilled workers (relative to unskilled ones) varied greatly.

In terms of our model, the main difference between the two scenarios is that the AFMP Shock is more concentrated regionally, thus leading to a higher value of Λ_{ms} than the Diaspora Shock. This means that the comparison of the two shocks will be instructive for evaluating the role of regional heterogeneity in terms of migrants.

Next, we explain how we parametrize the two versions of the model, then we present the results.

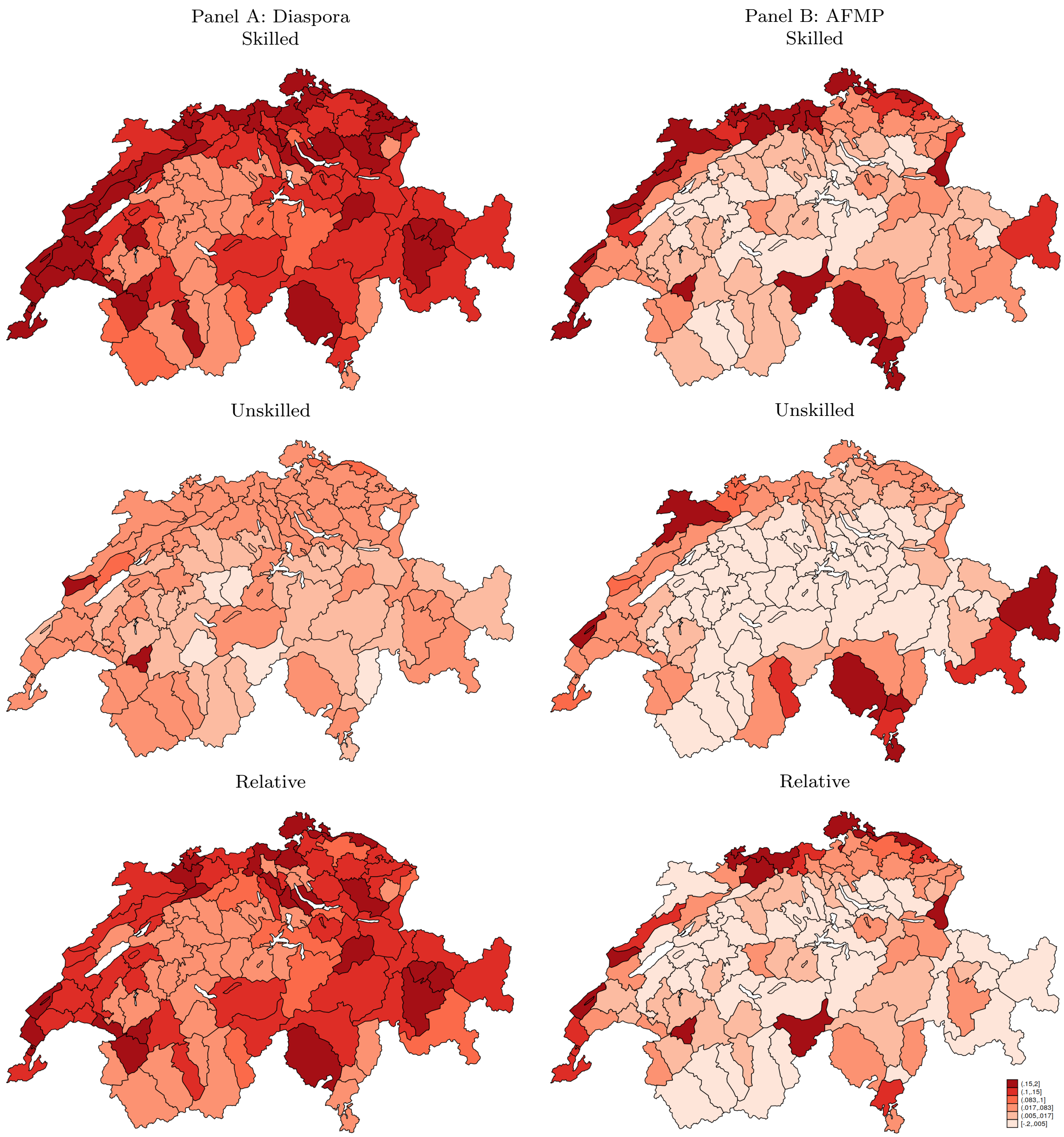
4.1 Benchmark Model

In the benchmark model, we need to quantify two types of parameters: (1) the elasticities of substitution δ and ε, and (2) the share parameters listed in Table 3, which capture the heterogeneity of firms and regions. Our data allow us to estimate δ, the elasticity of substitution between skilled and unskilled workers at the firm level, and to calibrate the share parameters, whereas we take the elasticity of substitution across products, ε, from the literature. In the following, we present the empirical strategy for the identification of δ and discuss briefly the quantification of the other parameters, before turning to the results of the simulations.

4.1.1 Estimation of δ

In this subsection, we estimate the firm-level elasticity of substitution between skilled and unskilled labor, δ, which plays a crucial role in the model. While the literature provides different values of the elasticity of substitution at the aggregate level (e.g., Katz and Murphy, 1992; Borjas, 2003;

Figure 4: Labor Supply Shocks



Notes: This figure depicts the two shocks that we bring to the simulations. Panel A shows the Diaspora Shock based on the shift-share for skilled, unskilled, and their ratio. Panel B shows the same for the AFMP Shock. Data Source: SESS.

D'Amuri et al., 2010; Felbermayr et al., 2010; Acemoglu and Autor, 2011; Brücker and Jahn, 2011; Manacorda et al., 2012; Ottaviano and Peri, 2012), our paper is the first, together with Bøler (2015), to provide a firm-level value for this elasticity. From firm i's demand for labor of skill e in equation (3), we obtain the following relative labor demand:

$$\ln\left(\frac{L_{si}}{L_{ui}}\right) = \lambda_i - \delta \ln\left(\frac{w_{sr}}{w_{ur}}\right), \qquad \lambda_i = \delta \ln\left(\frac{b_{si}}{b_{ui}}\right)$$

Note that this equation is the same for the extended version of the model, because the human-capital externality is skill-neutral at the firm level. Adding subscript t to indicate time, we can write the econometric counterpart of this equation as follows:

$$\ln\left(\frac{L^s_{irt}}{L^u_{irt}}\right) = -\delta \ln\left(\frac{w^s_{rt}}{w^u_{rt}}\right) + \underbrace{\kappa_r + \kappa_t + \kappa_j}_{\lambda_{i(j)rt}} + u_{irt} \tag{31}$$

where $\kappa_r, \kappa_t, \kappa_j$ denote region, time, and sector fixed effects. The assumption is that the relative efficiency of skilled and unskilled labor, as captured by $\lambda_{i(j)rt}$, can vary over time and between sectors and regions.[20] To correct for the possible correlation of the error component across firms within the same MS-region–sector, we cluster standard errors accordingly.[21]

To control for heterogeneous characteristics of workers such as experience or gender that are not embedded in our model but are present in the data, we use skill-specific wage rates, w^s_{rt} and w^u_{rt}, obtained from Mincerian regressions run separately by skill, as in Card (2009). These regressions include as controls a set of dummies for experience, detailed education categories, gender, and origin (Swiss or foreign) of the worker. Following Card (2009), we also account for heterogeneity in the labor demand for unskilled workers L_{ui}. As primary-educated workers earn 25% less than secondary-educated workers with similar characteristics, we assume that this wage difference is due to lower productivity. Thus we define L_{ui} as the employment of secondary-educated workers, plus 0.75 times the employment of primary-educated workers.

To properly identify δ, we need to acknowledge that any unobserved labor demand shock can simultaneously affect relative employment at the firm level and relative wages at the regional level. To solve this issue, we follow the shift-share approach of Card (2001) and exploit the tendency of new waves of immigrants to settle in areas where communities of the same nationality already exist.[22] Indeed, they can benefit from information about the location and can assimilate more easily into the host country. We exploit this idea to impute the number of immigrants arriving in each MS-region r and year t. This instrument provides an exogenous source of variation in the relative labor supply of workers at the MS-region level.

Since the SESS does not contain information on the workers' nationality, we combine data from

[20] Our data do not allow us to include firm fixed effects, as we are not able to follow firms over time.

[21] We restrict the estimation sample to firms employing both skilled and unskilled workers.

[22] Note that the AFMP Shock described before cannot be used to estimate δ, because many MS-regions contain municipalities that are in both the treated group and the control group, thus making identification difficult. However, the exercise can safely be used for predicting a simulation scenario that mimics the AFMP.

the Swiss Labour Force Survey (SLFS, a yearly survey that started in 1991) with the Swiss Census of 1980. The sample size varies between 15,000 and 75,000 individuals per year. Since 2003, it increased to more than 40,000 individuals. Thus we restrict our use of these data to the 2004–2014 period and take a two-year average to increase the sample size. We allocate the total number of recent immigrant workers from country c, level of education e, and year t, $L_{c,t}^{e,m}$, taken from the SLFS, in proportion to the share of that nationality group in the MS-region r as of 1980, calculated using the Swiss Census. We then sum across countries of origin c to obtain the imputed number of immigrants by MS-region r, education e, and year t:

$$L_{r,t}^{e,m,IV} = \sum_{c} \frac{L_{c,r,1980}^{m}}{L_{c,1980}^{m}} L_{c,t}^{e,m} \quad \text{where } e = s, u \quad t = \{2004, \ldots, 2014\}$$

Thus, the instrument for relative labor demand is defined as

$$\ln \left(\frac{L_{rt}^{s,m,IV}}{L_{rt}^{u,m,IV}} \right) \tag{32}$$

The use of weights that are far from the period of analysis is a great advantage that helps dissipate the concerns regarding this type of instrumental strategy (e.g., Jaeger et al., 2018). Specifically, dynamic adjustments due to the positive (or negative) effect of past settlements can hardly be a confounding factor for shocks that happened more than 20 years later. In 1980, the majority of immigrants were unskilled and employed in agriculture, construction, and hospitality as guest workers. In contrast, recent inflows of migrants are mainly composed of skilled workers employed in either manufacturing or high-tech service activities. Moreover, the distribution of immigrants in 1980 predates both the start of the SESS and the 2002 introduction of the AFMP, which significantly changed immigration patterns in Switzerland (Beerli et al., 2021; Ariu, 2022). Therefore, contemporaneous unobserved labor demand shocks are likely to be different and uncorrelated with those in 1980.

Panel A of Table 4 presents the second-stage estimates of the elasticity of substitution between skilled and unskilled labor at the firm level (δ). Column 1 reports the results for the complete sample: we find a negative and significant coefficient with a value around 2. This result remains the same when checking whether our estimate is driven by outliers. In column 2, we exclude all firms with fewer than five employees, and in column 3, we also exclude all firms in the top and bottom 2.5% of the distribution of firms' relative labor supply. As a further robustness check, we also run a more demanding regression that includes a richer set of fixed effects at the MS-region × sector level and the sector × year level.[23] Column 4 of Table 4 shows that the results remain similar and yield a value of δ in line with the main specification. In column 5 of Table 4, we use weights for 1970 to further dissipate issues related to dynamic adjustments that could be correlated with current unobserved shocks. Our estimates of the elasticity of substitution between skilled and unskilled

[23]In our setting, we cannot include the interaction between region and time fixed effects, since relative wages are defined at that level.

Table 4: Estimation of the Elasticity of Substitution Between Skill Groups ($-\delta$), 2004–2014

Dependent Variable: $\ln\left(\frac{L^s_{irt}}{L^u_{irt}}\right)$	(1)	(2)	(3)	(4)	(5)
Panel A: Second Stage					
$\ln\left(\frac{w^s_{rt}}{w^u_{rt}}\right)$	-1.999^c	-1.909^c	-2.240^b	-1.958^b	-2.268^b
	(1.034)	(1.079)	(0.956)	(0.886)	(1.064)
MS-Region FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
MS-Region x Sector FE	No	No	No	Yes	Yes
Sector x Year FE	No	No	No	Yes	Yes
N	141,246	128,962	122,587	122,587	122,587
R^2	0.25	0.26	0.25	0.25	0.25
Kleibergen-Paap Wald F	123.40	112.48	114.06	115.49	86.02
Panel B: First Stage					
$\ln\left(\frac{L^{s,m,IV}_{rt}}{L^{u,m,IV}_{rt}}\right)$	0.077^a	0.077^a	0.076^a	0.077^a	0.059^a
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)

Notes: Panel A reports the results of the second stage and Panel B of the first stage. In column 1, we present the results for the complete sample. In column 2, we drop firms with fewer than 5 employees. In column 3 to 5, we also drop the top and bottom 2.5% of the distribution of the log relative employment of firms. In column 5, we use the weights of the year 1970 to construct the instrument. Standard errors clustered at the MS-region–Sector level are reported in parenthesis. Significance levels at 10% (c), 5% (b) and 1% (a). Data source: SESS and SLFS

labor are slightly higher than the one estimated by Bøler (2015), i.e., 1.35. In our simulations, we will use a value of δ equal to 2.

We report the results of the first-stage regression in Panel B of Table 4. The first stage of the IV-estimation yields a positive sign. At first glance, it might seem surprising that an exogenous increase of the relative supply of skilled versus unskilled labor produces an increase in relative wages. Yet this result is consistent with the evidence presented in Beerli et al. (2021), who find that a skill-intensive immigration shock led to a stronger increase in skilled wages than in unskilled wages in Switzerland. These results are consistent with the human-capital-externalities version of our model, analyzed in Section 3. To complement this reduced-form evidence with a general-equilibrium quantification, we also run additional simulations that take into account these externalities in Section 4.2.

Recent developments in the use of the shift-share IV approach suggest how to test the quality of the instrumental strategy (e.g., Adão et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022). Our paper suffers less from the problems raised in these contributions, for two reasons. First, the unit of analysis is at a more disaggregated level than the endogenous variable. In other words, the shift-share constructed at the regional level should not capture employment dynamics at the firm level, given the large number of firms present in each one. Second, our instrument is constructed as the ratio of two instruments: the increase in the number of skilled workers in the numerator and the increase in the number of unskilled workers in the denominator. In this way, any omitted variable that symmetrically affects the inflows of both skilled and unskilled workers should cancel out. This last difference makes the use of the standard error correction suggested by Adão et al. (2019) inappropriate and the tests suggested by Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022) not directly applicable.[24]

4.1.2 Quantification of Other Parameters

The first parameter that needs to be identified is the elasticity of substitution across products, ε. Anderson and van Wincoop (2003) argue that studies at the sectoral level set its value to be between 5 and 10. Subsequent studies tend to find values within this range (e.g., Romalis, 2007; Costinot et al., 2012; Caliendo and Parro, 2015). We therefore use a value—8—that is somehow in between that range and exactly the same as the one estimated by Egger et al. (2012).

For all the other parameters, we can instead rely on our data. In particular, all share parameters contained in Table 3 are calibrated from the SESS data for 2004, the base year for our simulations. To compute cost shares and employment shares of skilled workers at the firm and regional levels, we aggregate the individual observations for workers using the appropriate weights. The final sample that we bring to the simulations includes 54,512 firms.

Having all the parameters at hand, we can compute the values of Ψ_r (i.e., the degree of heterogeneity between firms in region r, as stated by Proposition 1). We show its distribution in Figure A.1 in the Appendix. The average is 0.34—and we appreciate that there is wide variation in the degree of heterogeneity across regions. Then, we compute the aggregate elasticity of substitution between skilled and unskilled labor at the regional level (σ_r) and at the national level (σ). Our estimate of the latter yields a value of 4.3, which is in line with the estimates (at the aggregate level) found in the literature for European countries.[25]

4.1.3 Simulation Results

Our main quantitative findings regarding the mechanisms of the benchmark model are as follows. First, reallocation across firms within regions is the quantitatively most important mechanism in the benchmark version of the model. The change in the output mix within a region reduces the wage impact of immigration by more than half. Second, the heterogeneity in the geographical settlement of migrants can have a quantitatively significant effect, but its importance depends on the context of the migration shock. In our case, we find that the AFMP Shock (which is limited to border regions of Switzerland) reduces the negative wage impact for natives by almost one fifth, but the Diaspora Shock (which is geographically less concentrated) has a smaller impact. This implies that the apparent (aggregate) elasticity of substitution between natives and migrants of a given skill varies depending on the context. Third, the role of the mobility of workers between regions is relatively limited. Under the extreme assumption of perfect mobility, the "sheltering effect of the geographical segregation between migrants and natives disappears completely, but the reallocation

[24] To dissipate any doubt, in Appendix A.9 we split the numerator and denominator components of the instrument and test their robustness following Borusyak et al. (2022).

[25] For Germany, estimates of the (aggregate) elasticity of substitution between skilled and unskilled labor by D'Amuri et al. (2010), Felbermayr et al. (2010), and Brcker et al. (2014) are, respectively, 3, 4.6, and 4.4, while Manacorda et al. (2012) and Brcker et al. (2014) obtain estimates of, respectively, 5 and 4 for the UK, and Edo and Toubal (2015) estimate an elasticity of 4 for France. For the US, estimated values are generally lower: Borjas (2003), and Ottaviano and Peri (2012) find, respectively, values of 1.3 and 2.

effect across firms still operates fully. However, as argued in Section 2, the empirically relevant case is closer to the “no mobility assumption. We now turn to describing these results in detail.

We start by simulating a benchmark scenario in which we set the heterogeneity across regions (i.e., both Υ_s and Λ_{ms}) and across firms (i.e., Ψ_r) to zero and subsequently switch them on one by one. In the initial simulation, we assume that all firms have the same skill intensity, which is set to the national average ($s_{si} = \theta_s$). This implies that both the skill heterogeneity within regions, Ψ_r, and heterogeneity between regions, Υ_s, are zero. Moreover, we initially set the effects due to spatial segregation between natives and migrants (Λ_{ms}) to zero by assuming that the spatial distribution of “old” migrants across regions is uniform ($\theta_{msr} = \theta_{ms}$).[26]

Table 5 presents the percentage change in wages, relative wages, and implied substitution elasticities between migrants and natives (σ_{mn}) for the four different categories of workers in the case in which we consider the effect of an increase in the number of skilled workers (Panel A) or unskilled workers (Panel B) due to immigration that spreads in the different Swiss regions following preexisting diasporas. Column 1 of Table 5 shows that in the absence of firm (i.e., $\Psi_r = 0$) and regional heterogeneity (i.e., $\Upsilon_s = 0$ and $\Lambda_{ms} = 0$), skilled immigration decreases wages of skilled natives by 4.77%. By adding heterogeneity in the settlement of migrants across labor markets (i.e., $\Lambda_{ms} > 0$) in column 2, the impact on skilled native wages decreases (in absolute value) by 0.17 percentage points. In column 3, we account for skill heterogeneity between regions (but not within, i.e., $\Upsilon_s > 0$) by setting firms' skill intensity to the regional average ($s_{si} = \theta_{sr}$). This leads to a decrease of the negative effect of additional 0.04 percentage points.

Therefore, regional heterogeneity decreases the impact of skilled immigration on skilled natives by about 4.4%; most of the attenuation effect is due to the heterogeneity in migrants' settlement. This suggests that in the Swiss context skill clustering at the regional level is quantitatively less important than migrant clustering for the impact of skilled immigration on wages of native skilled workers. Accordingly, the results for migrants' wages (w_{ms}) show that due to migrant clustering, their wages decrease further than in the benchmark scenario. In other words, as explained in corollary 3.2, the fact that migrants are clustered in certain regions makes them suffer more from additional immigration while partially sheltering natives from wage decreases. At the aggregate level, skilled migrants and skilled natives therefore appear as imperfect substitutes.

In column 4, we introduce firm heterogeneity within regions (i.e., $\Psi_r > 0$) by setting s_{si} to their observed values. In this case, the impact of skilled migration on native skilled wages decreases by more than half, meaning that reallocation across firms represents the most important of our mechanisms to mitigate the negative impact of immigration. This is also evident when looking at the implied elasticity of substitution between migrants and natives within a skill category, σ_{mn}^s and σ_{mn}^u: their value increases substantially as we introduce firm heterogeneity.[27]

[26]This means that $\text{Var}(\theta_{msr}) = 0$, which implies that $\text{Cov}(\theta_{msr}, \theta^*_{msr}) = 0$, even if new migrants are not uniformly distributed across regions.

[27]Note that these substitution elasticities in simulations (1) and (5) are infinite, since migrants and natives have identical wages for a given skill. In simulation (1), this is because the model is equivalent to an aggregate production function at the national level, whereas in simulation (5) perfect mobility between regions equalizes wages for natives and migrants of a same skill.

Finally, in column 5, we check whether our results are due to the hypothesis of segmented labor markets by assuming that people can freely move across regions, and so Switzerland has a unique labor market. In this case, the impact of skilled immigrants is slightly more negative than in simulation (4) because labor mobility nullifies regional reallocations. Therefore, our results are not driven by the assumption of segmented labor markets.

When looking at the unskilled immigration shock in column 1 of Panel B of Table 5, its negative effect in absence of across-firm and across-region reallocation is substantially smaller than for the skilled case. As Figure 4 shows, this is because the influx of unskilled workers is milder than for the skilled case. When introducing heterogeneity across regions (columns 2–3) and across firms (column 4), results mimic quite closely the qualitative findings of panel A: the effect of unskilled immigration on native skilled wages is cut by half when both mechanisms are at play. However, in quantitative terms, the regional heterogeneity in terms of migrants accounts for a decrease in the negative wage effect by 10%, which is substantially higher than for skilled workers. This is because unskilled migrants tend to cluster more than high-skilled migrants (i.e., $\Lambda_{mu} > \Lambda_{ms}$). For the same reason, the negative effect on unskilled migrants in percentage terms is more important for the unskilled than for the skilled (10% against 18%). In other words, unskilled immigration shocks have milder negative effects on unskilled natives but stronger negative effects on unskilled migrants because unskilled migrants tend to be more clustered than skilled ones, which protects natives' wages but decreases those of immigrants more.

We present the results for the AFMP shock in Table 6. As explained before, this shock is milder than the Diaspora Shock. For this reason, Column 1 of panel A shows that the effect of a skilled immigration shock would decrease skilled native wages by 2.36% in the absence of the model's reallocation mechanisms. However, the decrease in the negative effect due to migrant heterogeneity across regions is more important than for the Diaspora Shock, decreasing the wage effect by 0.46 percentage points in column 2. This is because this shock is much more geographically concentrated in the regions in which there are already skilled migrants than the Diaspora Shock, thus strengthening the attenuation mechanism.

The effect of skill heterogeneity in column 3 is instead exactly the same as for the diaspora shock because their settlement does not change in the two shocks. As for the heterogeneity in terms of firm employment composition, column 4 shows that also in this context it cuts the negative effect of immigration by more than half. Finally, removing the assumption of segmented labor markets in column 5 increases the negative effect on average wages of skilled natives, since the sheltering effect of migrants' regional segregation does not operate anymore.

Switching to panel B of Table 6, we can evaluate the AFMP Shock in the case of an unskilled immigration shock in isolation. Also in this case, the effect of the shock is milder for unskilled than for skilled natives because unskilled inflows are milder also in this scenario. Moreover, the attenuation effect of Λ_{mu} in column 2 is stronger than for the skilled due to the higher concentration of unskilled workers across labor markets, thus implying a lower decrease in the wage of unskilled natives and a stronger decrease of unskilled immigrants' wages. The effect of Υ_s in column 3 is

Table 5: Diaspora Shock Simulation Results

	(1)	(2)	(3)	(4)	(5)
$\Psi_r > 0$	No	No	No	Yes	Yes
$\Upsilon_s > 0$	No	No	Yes	Yes	Yes
$\Lambda_{ms} > 0$	No	Yes	Yes	Yes	Yes
Across-region mobility	No	No	No	No	Yes
Panel A: Skilled Shock (unskilled = 0)					
w_{ns}	-4.77%	-4.60%	-4.56%	-2.19%	-2.21%
w_{nu}	2.14%	2.08%	1.97%	0.96%	0.96%
w_{ms}	-4.77%	-5.30%	-5.14%	-2.46%	-2.21%
w_{mu}	2.14%	2.28%	2.20%	1.01%	0.96%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.49%	-0.48%	-0.23%	-0.23%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.55%	-0.53%	-0.25%	-0.23%
σ^s_{mn}	∞	19.44	23.96	49.85	∞
Panel B: Unskilled Shock (skilled = 0)					
w_{ns}	1.23%	1.18%	1.13%	0.54%	0.53%
w_{nu}	-0.49%	-0.45%	-0.46%	-0.21%	-0.23%
w_{ms}	1.23%	1.38%	1.29%	0.57%	0.53%
w_{mu}	-0.49%	-0.58%	-0.60%	-0.29%	-0.23%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.52%	-0.50%	-0.48%	-0.23%	-0.23%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.52%	-0.59%	-0.58%	-0.26%	-0.23%
σ^u_{mn}	∞	25.93	23.48	39.84	∞

Notes: This table presents the simulation results for the Diaspora Shock. Each column presents a different scenario that can be used to evaluate the different mechanisms of the model.

exactly the same as in panel A, because the skill heterogeneity is the same in both cases, and the effect of Ψ_r in column 4 cuts the negative effect on unskilled by more than half.

In summary, the AFMP Shock has milder negative effects, but it also shows that regional reallocations can be quantitatively important in the presence of high immigrant segregation. This is especially highlighted by the values of σ^s_{mn} and σ^u_{mn}, which are lower for the AFMP Shock than for the Diaspora Shock due to the stronger sheltering effect of migration clustering across regional labor markets. In other words, the aggregate elasticity of substitution between migrants and natives varies depending on the degree to which migrants cluster geographically.

The results of the two shocks highlight that the mechanisms arising from our model are quantitatively important in understanding why the impact of immigration on wages is limited and why immigrants appear as imperfect substitutes at the aggregate level. Across-firm and across-region reallocations dampen the aggregate wage impact of immigration and make natives and migrants appear as imperfect substitutes at the aggregate level.

4.2 Simulation of the Extended Model With Human-Capital Externalities

To simulate the model with human-capital externalities and perform the same counterfactual scenarios, we first need to identify the relevant moments of the distribution of η_i.

Table 6: AFMP Shock Simulation Results

	(1)	(2)	(3)	(4)	(5)
$\Psi_r > 0$	No	No	No	Yes	Yes
$\Upsilon_s > 0$	No	No	Yes	Yes	Yes
$\Lambda_{ms} > 0$	No	Yes	Yes	Yes	Yes
Across-region mobility	No	No	No	No	Yes
Panel A: Skilled Shock (unskilled = 0)					
w_{ns}	-2.36%	-1.90%	-1.86%	-0.87%	-1.09%
w_{nu}	1.01%	0.87%	0.86%	0.44%	0.47%
w_{ms}	-2.36%	-3.75%	-3.61%	-1.79%	-1.09%
w_{mu}	1.01%	1.30%	1.29%	0.56%	0.47%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.41%	-0.40%	-0.19%	-0.23%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.75%	-0.73%	-0.35%	-0.23%
σ^s_{mn}	∞	3.66	3.85	7.30	∞
Panel B: Unskilled Shock (skilled = 0)					
w_{ns}	1.28%	1.13%	1.11%	0.55%	0.60%
w_{nu}	-0.56%	-0.39%	-0.38%	-0.15%	-0.26%
w_{ms}	1.28%	1.71%	1.65%	0.70%	0.60%
w_{mu}	-0.56%	-0.92%	-0.90%	-0.49%	-0.26%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.41%	-0.40%	-0.19%	-0.23%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	-0.50%	-0.71%	-0.69%	-0.32%	-0.23%
σ^u_{mn}	∞	6.95	7.09	10.69	∞

Notes: This table presents the simulation results for the AFMP Shock. Each column presents a different scenario that can be used to evaluate the different mechanisms of the model.

4.2.1 Estimation of the Distribution of η_i

The parameter η_i is unobserved, but we have two pieces of information that enable us to identify two moments of the distribution of η_i. First, the first stage of the IV estimation of δ provides an estimate of the elasticity of $\left(\frac{w_{sr}}{w_{ur}}\right)$ with respect to $\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$. Using proposition 1", we can then identify the (normalized) covariance Φ between η_i and s_{si}. Second, Moretti (2004) estimates the elasticity of firms' productivity with respect to the share of skilled workers in the region. This estimate can be related to the average of the η_i in our model.

To pin down the distribution of η_i, we have to rely on assumptions that allow us to use the sparse information at our disposal in a meaningful way. Therefore, we assume (1) that η_i can take only two values, 0 and η^*, and (2) that the probability that η_i takes the value η^* depends on firm i's skill intensity s_{si}. Specifically, we assume that this probability can be described by a logistic distribution:

$$Prob(\eta_i = \eta^* | s_{si}) = Prob\left(\eta_i/\eta^* = 1 | s_{si}\right) = \Lambda(\alpha + \beta s_{si}),$$

where $\Lambda(\cdot)$ is the cumulative distribution function of the logit model. This specification allows for a positive or negative correlation between s_{si} and η_i, depending on the sign of β.

To estimate the parameters α and β, we rely on the maximum likelihood estimators for the logit model. In this context, it is useful to redefine the dependent variable in the logit model as (η_i/η^*), which takes the values 0 or 1. The maximum likelihood estimators for α and β are given by the two following conditions (e.g., Greene, 2012, Ch. 17.3, equations 17–18),:

$$\sum_i s_i \left(\frac{\eta_i}{\eta^*} - \Lambda_i \right) = 0 \tag{33}$$

$$\sum_i s_i \left(\frac{\eta_i}{\eta^*} - \Lambda_i \right) s_{si} = 0 \tag{34}$$

where $\Lambda_i = \Lambda(\alpha + \beta s_{si})$. Equation (33) says that the average of predicted probabilities is equal to the proportion of nonzero η_i, which in turn is related to the average external effect of human capital, $\bar{\eta} = \sum_i s_i \eta_i$, as estimated by Moretti (2004). On the other hand, equation (34) speaks to the correlation Φ between η_i and s_{si}. This becomes clear by rearranging the two equations as follows:

$$\sum_i s_i \Lambda(\alpha + \beta s_{si}) = \frac{\bar{\eta}}{\eta^*}, \tag{35}$$

$$\sum_i s_i (s_{si} - \theta_s) \Lambda(\alpha + \beta s_{si}) = \frac{\Phi \theta_s (1 - \theta_s)}{\eta^*} \tag{36}$$

These two equations enable us to estimate α and β for given $\bar{\eta}$, η^*, Φ, and θ_s. First, we use proposition 1" to identify the (normalized) covariance Φ between η_i and s_{si}. In Table 4, the elasticity of (w_s/w_u) with respect to $\left(\bar{L}_s/\bar{L}_u\right)$ is estimated as 0.077, which yields $\Phi = 0.24$ from proposition 1".[28] Second, we set $\bar{\eta} = 0.11$, following Moretti (2004)'s estimate of the elasticity of firms' productivity with respect to the share of skilled workers in the region.[29] Third, we have no information on the dispersion of the η_i which would allow us to identify η^*. Therefore, we set $\eta^* = 0.66$, which means that on average one-sixth of the η_i are nonzero. We carry out sensitivity analyses to check the robustness of simulation results to this assumption. It turns out that the value of η^* has almost no influence on the different wage indicators in our simulations, as long as η^* exceeds a minimum value.[30]

Finally, solving equations (35) and (36) for α and β yields: $\alpha = -7.66$ and $\beta = 11.87$.

For the simulations, the elasticity η_i of a firm's output with respect to the regional externality

[28] For simplicity, we consider the country as one big region when we estimate the covariance between η_i and s_{si}. In this case, Proposition 1" can be rewritten as

$$\mathrm{d}\ln\left(\frac{w_s}{w_u}\right) = -\left(\frac{1 - (\varepsilon - 1)\lambda_u \Phi}{\sigma}\right) \mathrm{d}\ln\left(\frac{\bar{L}_s}{\bar{L}_u}\right), \qquad \Phi = \frac{\mathrm{Cov}(s_{si}, \eta_i)}{\theta_s(1-\theta_s)} = \sum_i s_i \frac{(s_{si} - \theta_s)(\eta_i - \bar{\eta})}{\theta_s(1-\theta_s)}, \qquad \bar{\eta} = \sum_i s_i \eta_i$$

$\lambda_u = \bar{L}_u/(\bar{L}_s + \bar{L}_u)$ and σ is defined in Proposition 1'. According to the first-stage estimates in Table 4, the elasticity of (w_s/w_u) with respect to $\left(\bar{L}_s/\bar{L}_u\right)$ is 0.077. Therefore, we can use the one-region version of Proposition 1" to estimate Φ as follows: $\Phi = (0.077\,\sigma + 1)/[(\varepsilon - 1)\lambda_u]$, which yields $\Phi = 0.24$.

[29] Moretti (2004)'s specification of the externality differs from ours. In our notation, his specification can be written as $\ln A_i = \gamma \bar{L}_{sr}/(\bar{L}_{sr} + \bar{L}_{ur})$. This implies that the elasticity of A_i with respect to the share of skilled workers in total employment of region r is equal to $\gamma \bar{L}_{sr}/(\bar{L}_{sr} + \bar{L}_{ur})$. His preferred estimates of γ range from 0.5 to 0.7, and $\bar{L}_{sr}/(\bar{L}_{sr} + \bar{L}_{ur})$ is between 0.16 and 0.19 in his sample, yielding a central estimate of 0.11 for $\bar{\eta}$.

[30] If η^* is too small for a given value of $\bar{\eta}$, the logit estimation (equations (35) and (36)) yields no solution for α and β. For $\bar{\eta} = 0.11$, this minimum value of η^* lies around 0.5. As long as we are beyond that threshold, our sensitivity analysis indicates that the choice of η^* hardly changes the simulation results.

is set to its expected value:

$$E(\eta_i) = \eta^* \Lambda_i = \eta^* \Lambda(\alpha + \beta s_{si}) = \eta^* \left(\frac{\exp(\alpha + \beta s_{si})}{1 + \exp(\alpha + \beta s_{si})} \right)$$

This relationship is shown in Figure A.2 in the Appendix.

4.2.2 Simulation Results

The extended model with human-capital externalities relies on two dimensions of firm heterogeneity: firms' skill intensity (s_{si}) and the elasticity of firm productivity with respect to the share of skilled workers in the region (η_i). The simulations will help us to evaluate the quantitative importance of these two dimensions of firm heterogeneity for each migration shock.[31] Column 1 in Tables 7 and 8 relies on the extended model with full heterogeneity with respect to s_{si} and η_i. In column 2, we shut down the heterogeneity with respect to η_i by assuming that all firms in a region r have the same elasticity of productivity (equal to the average elasticity in the region, $\bar{\eta}_r$). Column 3 assumes again heterogeneity with respect to η_i but eliminates the within-region heterogeneity in terms of firms' skill intensity (by assuming that s_{si} is equal to θ_{sr}). Finally, column 4 shuts down both dimensions of heterogeneity.

The simulation results confirm the theoretical results derived in Section 3.6. In the "fully heterogenous" case (column 1), skilled immigration leads to an increase in both skilled and unskilled wages and to an *increase* in relative (skilled to unskilled) wages (Panel A in Tables 7 and 8). This increase in relative wages is due to the aggregate skill bias of the human-capital externality, which arises only if both dimensions of heterogeneity interact, i.e., if a firm's skill intensity is positively correlated with η_i. The implied effect is sizeable, meaning that regional externalities can indeed explain why immigrants can positively affect the wages of natives.

Shutting down either of the two dimensions of heterogeneity eliminates this positive correlation and therefore eliminates the skill bias of the externality, as can be seen in columns 2 and 3. Without the heterogeneity in η_i in column 2, the external effect of regional human capital is neutral at the aggregate level and the relative wage effects of immigration are similar to those in the model without externality. Indeed, the relative wage effects in column 2 of Table 7 are close to those in column 4 of Table 5, but absolute wages are a little more than 1 percentage point higher, due to the externality (which is skill-neutral in this case).

In column 3 of Table 7, the assumption that firms have the same skill intensity within a region has two effects. First, it eliminates the skill bias of the externality, similar to column 2. Second, there is no change in the output mix, which absorbs part of the (regional) immigration shock. Therefore, aggregate elasticity of substitution between skilled and unskilled labor at the regional level is reduced to $\sigma_r = \delta$. For this reason, the relative wage effects in column 3 of Table 7 are similar to those in column 3 of Table 5 but, as in column 2, absolute wage levels are higher due to the externality.

[31] As in Section 3.6 above, we consider only the case without mobility of workers between regions.

Table 7: Diaspora Shock Simulation Results: Extended Model With Human-Capital Externalities

	(1)	(2)	(3)	(4)
Heterogeneity in s_{si}	Yes	Yes	No ($s_{si} = \theta_{sr}$)	No
Heterogeneity in η_i	Yes	No ($\eta_i = \bar{\eta}_r$)	Yes	No
Panel A: Skilled Shock (unskilled = 0)				
w_{ns}	1.58%	-1.02%	-3.40%	-3.40%
w_{nu}	0.91%	2.01%	3.02%	3.02%
w_{ms}	1.95%	-1.01%	-3.68%	-3.68%
w_{mu}	0.93%	2.24%	3.43%	3.43%
$d\ln\left(\frac{w_{ns}}{w_{nu}}\right) / d\ln\left(\frac{L_s}{L_u}\right)$	0.05%	-0.22%	-0.47%	-0.47%
$d\ln\left(\frac{w_{ms}}{w_{mu}}\right) / d\ln\left(\frac{L_s}{L_u}\right)$	0.07%	-0.24%	-0.52%	-0.52%
Panel B: Unskilled Shock (skilled = 0)				
w_{ns}	-0.45%	0.21%	0.80%	0.80%
w_{nu}	-0.20%	-0.48%	-0.74%	-0.74%
w_{ms}	-0.67%	0.14%	0.86%	0.86%
w_{mu}	-0.28%	-0.62%	-0.94%	-0.94%
$d\ln\left(\frac{w_{ns}}{w_{nu}}\right) / d\ln\left(\frac{L_s}{L_u}\right)$	0.08%	-0.21%	-0.47%	-0.47%
$d\ln\left(\frac{w_{ms}}{w_{mu}}\right) / d\ln\left(\frac{L_s}{L_u}\right)$	0.12%	-0.23%	-0.55%	-0.55%

Notes: This table presents the simulation results for the Diaspora Shock for the model with human-capital externalities. Each column presents a different scenario that can be used to evaluate the different mechanisms of the model.

Finally, eliminating the heterogeneity in both parameters (η_i and s_{si}) yields the same results as in column 3, because the heterogeneity with respect to η_i has no effect if all firms within a region have the same skill intensity.

Two other observations stand out among the results of the extended model. First, it is remarkable that the wages of skilled migrants increase even more than those of skilled natives (column 1). This can be explained by the fact that skilled migrants tend to work more often in firms with high skill intensity than skilled natives, thus skilled migrants benefit more from the positive effect of the externality. Second, the results for the unskilled shock (Panel B in Table 7 and Table 8) are the mirror image of the skilled shock. Here the human-capital externality is negative for skill-intensive firms and drives down wages of skilled workers. The effect on unskilled wages is almost identical with or without the externality (compare column 1 in Table 7 with column 4 in Table 5 and column 1 in Table 8 with column 4 in Table 6) but skilled workers experience a decrease in their wages, which indicates that the externality prevails over the complementarity effect.

5 Conclusion

We argue in this paper that an increase in labor supply due to immigration does not automatically translate into lower wages, for three reasons. First, skill-intensive firms and skill-abundant regions benefit from a skilled-labor supply shock and get to produce a larger share of national output,

Table 8: AFMP Shock Simulation Results: Extended Model With Human-Capital Externalities

	(1)	(2)	(3)	(4)
Heterogeneity in s_{si}	Yes	Yes	No ($s_{si} = \theta_{sr}$)	No
Heterogeneity in η_i	Yes	No ($\eta_i = \bar{\eta}_r$)	Yes	No
Panel A: Skilled Shock (unskilled = 0)				
w_{ns}	0.74%	-0.34%	-1.33%	-1.33%
w_{nu}	0.43%	0.89%	1.31%	1.31%
w_{ms}	1.17%	-0.82%	-2.65%	-2.65%
w_{mu}	0.50%	1.30%	2.04%	2.04%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	0.05%	-0.18%	-0.39%	-0.39%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	0.10%	-0.31%	-0.69%	-0.69%
Panel B: Unskilled Shock (skilled = 0)				
w_{ns}	-0.35%	0.26%	0.82%	0.82%
w_{nu}	-0.14%	-0.39%	-0.62%	-0.62%
w_{ms}	-0.85%	0.19%	1.14%	1.14%
w_{mu}	-0.45%	-0.90%	-1.31%	-1.31%
$\mathrm{d}\ln\left(\frac{w_{ns}}{w_{nu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	0.06%	-0.18%	-0.39%	-0.39%
$\mathrm{d}\ln\left(\frac{w_{ms}}{w_{mu}}\right) / \mathrm{d}\ln\left(\frac{L_s}{L_u}\right)$	0.11%	-0.30%	-0.66%	-0.66%

Notes: This table presents the simulation results for the AFMP Shock for the model with human-capital externalities. Each column presents a different scenario that can be used to evaluate the different mechanisms of the model.

thus reducing the extent of relative wage adjustments. Second, natives' average national wage is protected from immigration due to the uneven settlement of migrants across labor markets. Third, in the presence of human-capital externalities that are neutral at the firm level, but skill-biased on aggregate, skilled immigration can benefit skill-intensive firms and increase absolute and relative skilled wages. Our simulations find that these mechanisms are quantitatively strong and meaningful. Therefore, firm and regional heterogeneity, labor mobility, and human-capital externalities are compelling factors for understanding the wage consequences of the immigration phenomenon.

References

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Adão, R., Kolesär, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Ariu, A. (2022). Foreign Workers, Product Quality and Trade: Evidence from a Natural Experiment. *Journal of International Economics*, 139.

Baqaee, D. and Farhi, E. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, 87(4):1155–1203.

Beerli, A., Ruffner, J., Siegenthaler, M., and Peri, G. (2021). The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland. *American Economic Review*, 111(3):976–1012.

Beine, M., Docquier, F., and Ozden, C. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.

Blackorby, C. and Russell, R. R. (1989). Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities). *The American Economic Review*, 79(4):882–888.

Bøler, E. A. (2015). Technology-Skill Complementarity in a Globalized World. Mimeo.

Borjas, G. J. (2003). The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *Quarterly Journal of Economics*, 118(4):1335–1374.

Borjas, G. J. (2006). Native internal migration and the labor market impact of immigration. *The Journal of Human Resources*, 41(2):221–258.

Borjas, G. J. and Katz, L. F. (2007). The Evolution of the Mexican-Born Workforce in the United States. In *Mexican Immigration to the United States*, NBER Chapters, pages 13–56. National Bureau of Economic Research, Inc.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies*, 89(1):181–213.

Brinatti, A. and Morales, N. (2021). Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments. Working Paper 21-16, Federal Reserve Bank of Richmond.

Brücker, H. and Jahn, E. J. (2011). Migration and wage-setting: Reassessing the labor market effects of migration. *The Scandinavian Journal of Economics*, 113(2):286–317.

Brcker, H., Hauptmann, A., Jahn, E. J., and Upward, R. (2014). Migration and Imperfect Labor Markets: Theory and Cross-Country Evidence from Denmark, Germany and the UK. *European Economic Review*, 66:205–225.

Burchardi, K. B., Chaney, T., and Hassan, T. A. (2019). Migrants, Ancestors, and Foreign Investments. *Review of Economic Studies*, 86(4):1448–86.

Burstein, A., Hanson, G., Tian, L., and Vogel, J. (2020). Tradability and the labor-market impact of immigration: Theory and evidence from the united states. *Econometrica*, 88(3):1071–1112.

Caliendo, L. and Parro, F. (2015). Estimates of the Trade and Welfare Effects of NAFTA. *Review of Economic Studies*, 82(1):1–44.

Card, D. (2001). Immigrant inflows, native outflows and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Card, D. (2009). Immigration and Inequality. *American Economic Review*, 99(2):1–21.

Card, D. (2012). Comment: The elusive search for negative wage impacts of immigration. *Journal of the European Economic Association*, 10(1):211–215.

Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Costinot, A., Donaldson, D., and Komunjer, I. (2012). What goods do countries trade? a quantitative exploration of ricardo's ideas. *The Review of Economic Studies*, 79(2):581–608.

Cristelli, G. and Lissoni, F. (2020). Free Movement of Inventors: Open-Border Policy and Innovation in Switzerland. Paper 104120, MPRA.

D'Amuri, F., Ottaviano, G. I. P., and Peri, G. (2010). The Labor Market Impact of Immigration in Western Germany in the 1990s. *European Economic Review*, 54(4):550–570.

Di Giovanni, J., Levchenko, A. A., and Ortega, F. (2015). A Global View of Cross-Border Migration. *Journal of the European Economic Association*, 13(1):168–202.

Dustmann, C., Frattini, T., and Preston, I. P. (2012). The effect of immigration along the distribution of wages. *The Review of Economic Studies*, 80(1):145–173.

Dustmann, C. and Glitz, A. (2015). How do Industries and Firms Respond to Changes in Local Labor Supply? *Journal of Labor Economics*, 33(3):711–750.

Edo, A. and Toubal, F. (2015). Selective immigration policies and wages inequality. *Review of International Economics*, 23(1):160–187.

Egger, P., Larch, M., and Staub, K. E. (2012). Trade preferences and bilateral trade in goods and services: A structural approach. Cepr discussion paper.

Emami Namini, J., Facchini, G., and López, R. A. (2015). A Model of Firm Heterogeneity in Factor Intensities and International Trade. *Canadian Journal of Economics*, 48(4):1456–1480.

Felbermayr, G., Geis, W., and Kohler, W. (2010). Restrictive Immigration Policy in Germany: Pains and Gains foregone? *Review of World Economics*, 146(1):1–21.

Ghosh, A., Mayda, A. M., and Ortega, F. (2014). The Impact of Skilled Foreign Workers on Firms: An Investigation of Publicly Traded U.S. Firms. Iza discussion paper.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Gray, R., Montresor, G., and Wright, G. C. (2020). Processing Immigration Shocks: Firm Responses on the Innovation Margin. *Journal of International Economics*, 126.

Greene, W. H. (2012). *Econometric Analysis*. Pearson Education, 7 edition.

Hornung, E. (2014). Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia. *American Economic Review*, 104(1):84–122.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-Share Instruments and the Impact of Immigration. Nber working paper.

Javorcik, B. S., Özden, Ç., Spatareanu, M., and Neagu, C. (2011). Migrant Networks and Foreign Direct Investment. *Journal of Development Economics*, 94(2):231–241.

Katz, L. and Murphy, K. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Kerr, W. R. and Lincoln, W. F. (2010). The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention. *Journal of Labor Economics*, 28(3):473–508.

Kugler, M. and Rapoport, H. (2007). International Labor and Capital Flows: Complements or Substitutes? *Economics Letters*, 94(2):155–162.

Llull, J. (2017). Immigration, Wages, and Education: A Labour Market Equilibrium Structural Model. *The Review of Economic Studies*, 85(3):1852–1896.

Losa, F., Bigotta, M., and Gonzalez, O. (2012). Libera Circolazione: Gioie o Dolori?: Valutazione degli Impatti sul Mercato del Lavoro Svizzero dell'Abolizione della priorità d'Impiego ai Lavoratori Indigeni. Technical report.

Manacorda, M., Manning, A., and Wadsworth, J. (2012). The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain. *Journal of the European Economic Association*, 10(1):120–151.

Mayda, A. M., Ortega, F., Peri, G., Shih, K., and Sparber, C. (2020). Coping with H-1B Shortages: Firm Performance and Mitigation Strategies. NBER Working Paper 27730.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Mitaritonna, C., Orefice, G., and Peri, G. (2017). Immigrants and Firms' Outcomes: Evidence from France. *European Economic Review*, 96:62–82.

Monras, J. (2018). Economic shocks and internal migration. CEPR Discussion Paper 12977.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, 94(3):656–690.

Oberfield, E. and Raval, D. (2021). Micro data and macro technology. *Econometrica*, 89(2):703–732.

Ottaviano, G. I. P. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Peri, G., Shih, K., and Sparber, C. (2015). STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics*, 33(S1):225–255.

Peri, G. and Sparber, C. (2009). Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics*, 1(3):135–69.

Peri, G. and Sparber, C. (2011). Assessing Inherent Model Bias: An Application to Native Displacement in Response to Immigration. *Journal of Urban Economics*, 69(1):82–91.

Romalis, J. (2007). Nafta's and cusfta's impact on international trade. *The Review of Economics and Statistics*, 89(3):416–435.

Sato, R. and Koizumi, T. (1973). On the Elasticities of Substitution and Complementarity. *Oxford economic papers*, 25(1):44–56.

A Appendix

A.1 Equations of the Model

The model is defined by equations (3), (4), and (5) in the main text, as well as by the following equations, which we omitted from the model presentation for the sake of simplicity:

- Conditional demand functions for intermediate goods:

$$y_i = \beta_i^{\varepsilon} \left(\frac{p_i}{P}\right)^{-\varepsilon} Y \tag{A.1}$$

- Dual price function of the final good's price P (which is the numéraire):

$$P = \left(\sum_{i \in \Omega} \beta_i^{\varepsilon} p_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}, \qquad P = 1 \tag{A.2}$$

- Markup and marginal cost equations:

$$p_i = \left(\frac{\varepsilon}{\varepsilon - 1}\right) c_i \tag{A.3}$$

$$c_i = \frac{1}{A_i} \left(b_{si}^{\delta} w_{sr}^{1-\delta} + b_{ui}^{\delta} w_{ur}^{1-\delta}\right)^{\frac{1}{1-\delta}}. \tag{A.4}$$

A.2 Differentiation of the model

Due to Walras' law, one of the equilibrium equations of the model is redundant (i.e., it is a linear combination of the other equations of the model), but it turns out to be convenient to differentiate all equations. Differentiating these equations with respect to all endogenous variables and the regional labor supplies $\bar{L}_{sr}$ and $\bar{L}_{ur}$ yields the following linear system:

- Demand equations:

$$\begin{aligned} \mathrm{d}\ln y_i &= -\varepsilon\, \mathrm{d}\ln c_i + \mathrm{d}\ln Y \\ \mathrm{d}\ln L_{si} &= \delta\, \mathrm{d}\ln c_i - \delta\, \mathrm{d}\ln w_{sr} + \mathrm{d}\ln y_i \\ \mathrm{d}\ln L_{ui} &= \delta\, \mathrm{d}\ln c_i - \delta\, \mathrm{d}\ln w_{ur} + \mathrm{d}\ln y_i \end{aligned}$$

- Price block:

$$\begin{aligned} 0 &= \sum_i s_i\, \mathrm{d}\ln c_i \\ \mathrm{d}\ln c_i &= s_{si}\, \mathrm{d}\ln w_{sr} + (1 - s_{si})\, \mathrm{d}\ln w_{ur} \end{aligned}$$

- Labor-market equilibrium and aggregate income:

$$\begin{aligned}
\sum_{i\in\Omega_r}\lambda_{sri}\,\mathrm{d}\ln L_{si} &= \mathrm{d}\ln\bar{L}_{sr}\\
\sum_{i\in\Omega_r}\lambda_{uri}\,\mathrm{d}\ln L_{ui} &= \mathrm{d}\ln\bar{L}_{ur}\\
\mathrm{d}\ln Y &= \sum_r s_r\left(\theta_{sr}\,\mathrm{d}\ln\bar{L}_{sr}+(1-\theta_{sr})\,\mathrm{d}\ln\bar{L}_{ur}\right)
\end{aligned}$$

A.3 Some Tedious but Useful Results (Propositions 1 and 2)

In this section, we give the details of the proof of Proposition 1 and some intermediate results that are also useful for Proposition 2. We have

$$\lambda_{sri}=\frac{s_{si}s_{ir}}{\theta_{sr}}\text{ and }\lambda_{uri}=\frac{(1-s_{si})s_{ir}}{1-\theta_{sr}},\text{ where }\theta_{sr}=\sum_{i\in\Omega_r}s_{si}s_{ir}$$

since

$$\lambda_{sri}=\frac{L_{si}}{\bar{L}_{sr}}=\frac{w_{sr}L_{si}}{w_{sr}\bar{L}_{sr}}=\underbrace{\frac{w_{sr}L_{si}}{w_{sr}L_{si}+w_{ur}L_{ui}}}_{s_{si}}\cdot\underbrace{\frac{w_{sr}L_{si}+w_{ur}L_{ui}}{w_{sr}\bar{L}_{sr}+w_{ur}\bar{L}_{ur}}}_{s_{ir}}\cdot\underbrace{\frac{w_{sr}\bar{L}_{sr}+w_{ur}\bar{L}_{ur}}{w_{sr}\bar{L}_{sr}}}_{1/\theta_{sr}}$$

and

$$\lambda_{uri}=\frac{L_{ui}}{\bar{L}_{ur}}=\frac{w_{ur}L_{ui}}{w_{ur}\bar{L}_{ur}}=\underbrace{\frac{w_{ur}L_{ui}}{w_{sr}L_{si}+w_{ur}L_{ui}}}_{1-s_{si}}\cdot\underbrace{\frac{w_{sr}L_{si}+w_{ur}L_{ui}}{w_{sr}\bar{L}_{sr}+w_{ur}\bar{L}_{ur}}}_{s_{ir}}\cdot\underbrace{\frac{w_{sr}\bar{L}_{sr}+w_{ur}\bar{L}_{ur}}{w_{ur}\bar{L}_{ur}}}_{1/(1-\theta_{sr})}$$

Therefore,

$$\sum_{i\in\Omega_r}s_{si}\lambda_{sri} = \sum_{i\in\Omega_r}\frac{s_{ir}s_{si}^2}{\theta_{sr}} \tag{A.5}$$

$$\sum_{i\in\Omega_r}s_{si}\lambda_{uri} = \sum_{i\in\Omega_r}\frac{s_{ir}(1-s_{si})s_{si}}{1-\theta_{sr}}=\frac{\theta_{sr}-\sum_{i\in\Omega_r}s_{ir}s_{si}^2}{1-\theta_{sr}} \tag{A.6}$$

Finally,

$$\begin{aligned}
\sum_{i\in\Omega_r}s_{si}(\lambda_{sri}-\lambda_{uri}) &= \sum_{i\in\Omega_r}\left(\frac{s_{ir}s_{si}^2}{\theta_{sr}}-\frac{s_{ir}(1-s_{si})s_{si}}{1-\theta_{sr}}\right)=\sum_{i\in\Omega_r}\frac{s_{ir}[s_{si}^2(1-\theta_{sr})-(1-s_{si})s_{si}\theta_{sr}]}{\theta_{sr}(1-\theta_{sr})}\\
&= \sum_{i\in\Omega_r}\frac{s_{ir}s_{si}(s_{si}-\theta_{sr})}{\theta_{sr}(1-\theta_{sr})}=\sum_{i\in\Omega_r}\frac{s_{ir}(s_{si}-\theta_{sr})^2}{\theta_{sr}(1-\theta_{sr})}
\end{aligned} \tag{A.7}$$

where the last equality in equation (A.7) uses the fact that $\sum_{i\in\Omega_r}s_{ir}(s_{si}-\theta_{sr})\theta_{sr}=0$, since $\sum_{i\in\Omega_r}s_{ir}s_{si}=\theta_{sr}$. This completes the proof of Proposition 1.

A.4 More Tedious but Nonetheless Useful Results (Proposition 3)

The average wage of skilled native workers is $w_{ns} = \sum_r w_{sr} N_{sr} / \sum_{r'} N_{sr'}$. The relative change in w_{ns} can therefore be written as

$$\frac{\mathrm{d}w_{ns}}{w_{ns}} = \sum_r \omega_{nsr} \frac{\mathrm{d}w_{sr}}{w_{sr}}$$

where

$$\omega_{nsr} = \frac{w_{sr} N_{sr}}{\sum_{r'} w_{sr'} N_{sr'}} = \underbrace{\frac{w_{sr} N_{sr}}{w_{sr} \bar{L}_{sr}}}_{1-\theta_{msr}} \underbrace{\frac{w_{sr} \bar{L}_{sr}}{\sum_{r'} w_{sr'} \bar{L}_{sr'}}}_{\alpha_{sr}} \underbrace{\frac{\sum_{r'} w_{sr'} \bar{L}_{sr'}}{\sum_{r'} w_{sr'} N_{sr'}}}_{1/(1-\theta_{ms})} = \frac{\alpha_{sr}(1-\theta_{msr})}{1-\theta_{ms}}$$

Proposition 3 can be obtained by combining the results in Proposition 2 with equations (20) and (21):

$$\begin{aligned} \frac{\mathrm{d}w_{ns}}{w_{ns}} &= \sum_r \frac{\alpha_{sr}(1-\theta_{msr})}{1-\theta_{ms}} \frac{\mathrm{d}w_{sr}}{w_{sr}} \\ &= \sum_r \frac{\alpha_{sr}(1-\theta_{msr})}{1-\theta_{ms}} \left[\frac{1}{\varepsilon} \left(\mathrm{d}\ln Y - \mathrm{d}\ln Y_r \right) + \frac{1}{\sigma_r} \left(\mathrm{d}\ln Y_r - \mathrm{d}\ln \bar{L}_{sr} \right) \right] \end{aligned} \tag{A.8}$$

We can now use the fact that $\sum_r \frac{\alpha_{sr}(1-\theta_{msr})}{1-\theta_{ms}} = 1$ and the following expressions (from Proposition 2, assuming $\mathrm{d}\ln \bar{L}_{ur} = 0$): $\mathrm{d}\ln Y = \sum_r s_r \theta_{sr} \,\mathrm{d}\ln \bar{L}_{sr}$, $\mathrm{d}\ln Y_r = \theta_{sr} \,\mathrm{d}\ln \bar{L}_{sr}$, and $\mathrm{d}\ln Y_r - \mathrm{d}\ln \bar{L}_{sr} = -(1-\theta_{sr}) \,\mathrm{d}\ln \bar{L}_{sr}$. Equation (A.8) can therefore be rewritten as

$$\begin{aligned} \frac{\mathrm{d}w_{ns}}{w_{ns}} &= \frac{1}{\varepsilon} \left(\sum_r s_r \theta_{sr} \,\mathrm{d}\ln \bar{L}_{sr} - \sum_r \frac{\alpha_{sr}(1-\theta_{msr})\theta_{sr}}{1-\theta_{ms}} \,\mathrm{d}\ln \bar{L}_{sr} \right) \\ &\quad - \sum_r \frac{1}{\sigma_r} \frac{\alpha_{sr}(1-\theta_{msr})(1-\theta_{sr})}{1-\theta_{ms}} \,\mathrm{d}\ln \bar{L}_{sr} \\ &= \sum_r \left(\frac{1}{\varepsilon} \frac{\alpha_{sr}\theta_s(1-\theta_{ms}) - \alpha_{sr}(1-\theta_{msr})\theta_{sr}}{1-\theta_{ms}} - \frac{1}{\sigma_r} \frac{\alpha_{sr}(1-\theta_{msr})(1-\theta_{sr})}{1-\theta_{ms}} \right) \mathrm{d}\ln \bar{L}_{sr} \end{aligned}$$

since $s_r \theta_{sr} = \alpha_{sr} \theta_s$. Using equation (20) to substitute for $\mathrm{d}\ln \bar{L}_{sr}$ yields

$$\begin{aligned} \frac{\mathrm{d}w_{ns}}{w_{ns}} &= \left(\frac{1}{\varepsilon} \sum_r \frac{\alpha_{sr}[\theta_s(1-\theta_{ms})\theta^*_{ms} - \theta_{sr}(1-\theta_{msr})\theta^*_{msr}]}{\theta_{ms}(1-\theta_{ms})} \right. \\ &\quad \left. - \sum_r \frac{1}{\sigma_r} \frac{\alpha_{sr}(1-\theta_{sr})(1-\theta_{msr})\theta^*_{msr}}{\theta_{ms}(1-\theta_{ms})} \right) \frac{\mathrm{d}M_s}{L_s}, \end{aligned}$$

This establishes Proposition 3.

For Corollary 3.1, we assume $\theta_{msr} = \theta_{ms}$, $\theta^*_{msr} = \theta^*_{ms}$, and $\sigma_r = \sigma$. Equation (22) can therefore be rewritten as

$$\frac{\mathrm{d}w_{ns}}{w_{ns}} = \left(\frac{1}{\varepsilon}\sum_r \frac{\alpha_{sr}(\theta_s-\theta_{sr})(1-\theta_{ms})\theta^*_{ms}}{\theta_{ms}(1-\theta_{ms})} - \frac{1}{\sigma}\sum_r \frac{\alpha_{sr}(1-\theta_{sr})(1-\theta_{ms})\theta^*_{ms}}{\theta_{ms}(1-\theta_{ms})}\right)\frac{\mathrm{d}M_s}{L_s},$$

Furthermore, we have

$$\sum_r \alpha_{sr}(\theta_s-\theta_{sr}) = \sum_r \frac{s_r\theta_{sr}(\theta_s-\theta_{sr})}{\theta_s} = -\sum_r \frac{s_r(\theta_{sr}-\theta_s)^2}{\theta_s} = -\frac{\mathrm{Var}(\theta_{sr})}{\theta_s}$$

since $\sum_r s_r\theta_s(\theta_{sr}-\theta_s)=0$, and

$$\sum_r \alpha_{sr}(1-\theta_{sr}) = 1-\frac{\sum_r s_r\theta^2_{sr}}{\theta_s} = 1-\frac{\sum_r s_r(\theta_{sr}-\theta_s)^2+\theta_s^2}{\theta_s} = (1-\theta_s)\left(1-\frac{\mathrm{Var}(\theta_{sr})}{\theta_s(1-\theta_s)}\right)$$

since $\sum_r s_r(\theta_{sr}-\theta_s)^2 = \sum_r s_r(\theta_{sr}-\theta_s)\theta_{sr} = \sum_r s_r\theta^2_{sr}-\theta^2_s$

Combining these elements yields

$$\mathrm{d}\ln w_{ns} = -(1-\theta_s)\left[\frac{1}{\varepsilon}\cdot\frac{\mathrm{Var}(\theta_{sr})}{\theta_s(1-\theta_s)}+\frac{1}{\sigma}\left(1-\frac{\mathrm{Var}(\theta_{sr})}{\theta_s(1-\theta_s)}\right)\right]\frac{\theta^*_{ms}}{\theta_{ms}}\frac{\mathrm{d}M_s}{L_s}$$

which establishes Corollary 3.1.

For Corollary 3.2, we assume $\theta_{sr}=\theta_s$ and $\sigma_r=\sigma$. Equation (22) can therefore be rewritten as

$$\begin{aligned}\frac{\mathrm{d}w_{ns}}{w_{ns}} &= \left(\frac{1}{\varepsilon}\sum_r \frac{\alpha_{sr}[(1-\theta_{ms})\theta^*_{ms}-(1-\theta_{msr})\theta^*_{msr}]\theta_s}{\theta_{ms}(1-\theta_{ms})}\right. \\ &\quad \left.-\frac{1}{\sigma}\sum_r \frac{\alpha_{sr}(1-\theta_s)(1-\theta_{msr})\theta^*_{msr}}{\theta_{ms}(1-\theta_{ms})}\right)\frac{\mathrm{d}M_s}{L_s}\end{aligned} \tag{A.9}$$

Using the fact that $\sum_r \alpha_{sr}\theta_{msr}=\theta_{ms}$ and $\sum_r \alpha_{sr}\theta^*_{msr}=\theta^*_{ms}$, we have

$$\mathrm{Cov}(\theta_{msr},\theta^*_{msr}) = \sum_r \alpha_{sr}(\theta_{msr}-\theta_{ms})(\theta^*_{msr}-\theta^*_{ms}) = \sum_r \alpha_{sr}(\theta_{msr}\theta^*_{msr}-\theta_{ms}\theta^*_{ms})$$

Therefore

$$\begin{aligned}\sum_r \alpha_{sr}[(1-\theta_{ms})\theta^*_{ms}-(1-\theta_{msr})\theta^*_{msr}] &= \sum_r \alpha_{sr}(\theta_{msr}\theta^*_{msr}-\theta_{ms}\theta^*_{ms}) \\ &= \mathrm{Cov}(\theta_{msr},\theta^*_{msr}),\end{aligned}$$

$$\begin{aligned}\sum_r \alpha_{sr}(1-\theta_{msr})\theta^*_{msr} &= \sum_r \alpha_{sr}(\theta^*_{ms}-\theta_{msr}\theta^*_{msr}) \\ &= \sum_r \alpha_{sr}(\theta^*_{ms}-\theta_{ms}\theta^*_{ms}+\theta_{ms}\theta^*_{ms}-\theta_{msr}\theta^*_{msr}) \\ &= \theta^*_{ms}(1-\theta_{ms})-\mathrm{Cov}(\theta_{msr},\theta^*_{msr})\end{aligned}$$

Substituting these two expressions into (A.9) yields

$$\frac{\mathrm{d}w_{ns}}{w_{ns}} = \left(\frac{\theta_s}{\varepsilon}\frac{\mathrm{Cov}(\theta_{msr},\theta^*_{msr})}{\theta_{ms}(1-\theta_{ms})} - \frac{(1-\theta_s)}{\sigma}\frac{[\theta^*_{ms}(1-\theta_{ms}) - \mathrm{Cov}(\theta_{msr},\theta^*_{msr})]}{\theta_{ms}(1-\theta_{ms})}\right)\frac{\mathrm{d}M_s}{L_s}$$

Collecting terms, we obtain Corollary 3.2:

$$\frac{\mathrm{d}w_{ns}}{w_{ns}} = \left[-\frac{(1-\theta_s)}{\sigma}\cdot\frac{\theta^*_{ms}}{\theta_{ms}} + \frac{\mathrm{Cov}(\theta_{msr},\theta^*_{msr})}{\theta_{ms}(1-\theta_{ms})}\left(\frac{\theta_s}{\varepsilon} + \frac{(1-\theta_s)}{\sigma}\right)\right]\frac{\mathrm{d}M_s}{L_s}$$

A.5 Differentiation of the Extended Model

Differentiating the equations of the extended model (with human-capital externalities) with respect to all endogenous variables and the regional labor supplies $\bar{L}_{sr}$ and $\bar{L}_{ur}$ yields the following linear system:

- Demand equations:

$$\begin{aligned}
\mathrm{d}\ln y_i &= -\varepsilon\,\mathrm{d}\ln c_i + \mathrm{d}\ln Y \\
\mathrm{d}\ln L_{si} &= \delta\,\mathrm{d}\ln c_i - \delta\,\mathrm{d}\ln w_{sr} + \mathrm{d}\ln y_i + (\delta-1)\,\mathrm{d}\ln A_i \\
\mathrm{d}\ln L_{ui} &= \delta\,\mathrm{d}\ln c_i - \delta\,\mathrm{d}\ln w_{ur} + \mathrm{d}\ln y_i + (\delta-1)\,\mathrm{d}\ln A_i
\end{aligned}$$

- Regional externality:

$$\mathrm{d}\ln A_i = \lambda_{ur}\eta_i(\mathrm{d}\ln\bar{L}_{sr} - \mathrm{d}\ln\bar{L}_{ur})$$

- Price block:

$$\begin{aligned}
0 &= \sum_i s_i\,\mathrm{d}\ln c_i \\
\mathrm{d}\ln c_i &= s_{si}\,\mathrm{d}\ln w_{sr} + (1-s_{si})\,\mathrm{d}\ln w_{ur} - \mathrm{d}\ln A_i
\end{aligned}$$

- Labor-market equilibrium and aggregate income:

$$\begin{aligned}
\sum_{i\in\Omega_r}\lambda_{sri}\,\mathrm{d}\ln L_{si} &= \mathrm{d}\ln\bar{L}_{sr} \\
\sum_{i\in\Omega_r}\lambda_{uri}\,\mathrm{d}\ln L_{ui} &= \mathrm{d}\ln\bar{L}_{ur} \\
\mathrm{d}\ln Y &= \sum_r s_r\left(\theta_{sr}\,\mathrm{d}\ln\bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln\bar{L}_{ur}\right)
\end{aligned}$$

As in the simple model, one of the equilibrium equations of the model is redundant due to Walras' law.

A.6 Extended Model With Human-Capital Externalities at the Regional Level (Proposition 1”)

To see how a skilled immigration shock affects relative wages at the regional level, we differentiate the conditional labor demand and marginal cost equations and use the equilibrium conditions on the regional labor markets. We can then follow the same steps as in the derivation of Proposition 1. Equations (6) and (7) are changed to

$$(\delta-\varepsilon)\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right)\sum_{i\in\Omega_r}s_{si}\lambda_{sri}+(\delta-\varepsilon)\,\mathrm{d}\ln w_{ur}-\delta\,\mathrm{d}\ln w_{sr}+\,\mathrm{d}\ln Y-(1-\varepsilon)\sum_{i\in\Omega_r}\lambda_{sri}\,\mathrm{d}\ln A_i = \,\mathrm{d}\ln\bar{L}_{sr}$$

$$(\delta-\varepsilon)\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right)\sum_{i\in\Omega_r}s_{si}\lambda_{uri}+(\delta-\varepsilon)\,\mathrm{d}\ln w_{ur}-\delta\,\mathrm{d}\ln w_{ur}+\,\mathrm{d}\ln Y-(1-\varepsilon)\sum_{i\in\Omega_r}\lambda_{uri}\,\mathrm{d}\ln A_i = \,\mathrm{d}\ln\bar{L}_{ur}$$

where the change in TFP is given by $\,\mathrm{d}\ln A_i = \lambda_{ur}\eta_i\,\mathrm{d}\ln(\bar{L}_{sr}/\bar{L}_{ur})$. To eliminate Y, take the difference between these two equations,

$$[\Psi_r(\delta-\varepsilon)-\delta]\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right)=[1-(\varepsilon-1)\lambda_{ur}\Phi_r]\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$$

where $\Phi_r=\sum_{i\in\Omega_r}(\lambda_{sri}-\lambda_{uri})\eta_i$. This leads to Proposition 1”.

A.7 Aggregate Representation of the Extended Model With Externalities

To establish the link between the parameters of the aggregate representation of the model with regional externalities and the underlying structural parameters, we calculate the impact of a labor supply shock on regional output, prices, and (relative) wages in both versions of the model. First, we derive the results for the disaggregated model with regional externalities, following the same steps as in Section 3.3. With regional human-capital externalities, equation (12) is modified as follows:

$$\mathrm{d}\ln Y_r=\theta_{sr}\,\mathrm{d}\ln\bar{L}_{sr}+(1-\theta_{sr})\,\mathrm{d}\ln\bar{L}_{ur}+\lambda_{ur}\bar{\eta}_r\,\mathrm{d}\ln(\bar{L}_{sr}/\bar{L}_{ur}) \tag{A.10}$$

and equation (14) can be written as

$$\mathrm{d}\ln P_r=\theta_{sr}\,\mathrm{d}\ln w_{sr}+(1-\theta_{sr})\,\mathrm{d}\ln w_{ur}-\lambda_{ur}\bar{\eta}_r\,\mathrm{d}\ln(\bar{L}_{sr}/\bar{L}_{ur}) \tag{A.11}$$

since $\bar{\eta}_r=\sum_{i\in\Omega_r}s_{ir}\eta_i$. Equation (15) is unchanged and thus we have

$$\mathrm{d}\ln P_r=\frac{1}{\varepsilon}(\,\mathrm{d}\ln Y_r-\,\mathrm{d}\ln Y)$$

Finally, combining the latter equation with (A.11) yields the equivalent of equation (16),

$$\mathrm{d}\ln w_{sr} = \frac{1}{\varepsilon}(\,\mathrm{d}\ln Y - \,\mathrm{d}\ln Y_r) + (1-\theta_{sr})\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) + \lambda_{ur}\bar{\eta}_r\,\mathrm{d}\ln(\bar{L}_{sr}/\bar{L}_{ur}) \tag{A.12}$$

We now turn to the aggregate representation of the model with regional externalities, which is described by the nested CES in Panel B of Figure 3, where the lower level of the nested CES production function is given by equation (30) in the main text. Differentiating equation (30) yields $\mathrm{d}\ln Y_r = \theta_{sr}\,\mathrm{d}\ln\bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln\bar{L}_{ur} + \mathrm{d}\ln A_r + \theta_{sr}\,\mathrm{d}\ln B_r$, where $\mathrm{d}\ln A_r = \lambda_{ur}\kappa_r\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$ and $\mathrm{d}\ln B_r = \lambda_{ur}\gamma_r\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$. Therefore

$$\mathrm{d}\ln Y_r = \theta_{sr}\,\mathrm{d}\ln\bar{L}_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln\bar{L}_{ur} + \lambda_{ur}\left(\kappa_r + \theta_{sr}\gamma_r\right)\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) \tag{A.13}$$

On the other hand, given the constant markup, the change in the regional price P_r is equal to the change of the regional unit cost. Differentiating equation (A.17) yields

$$\mathrm{d}\ln P_r = \,\mathrm{d}\ln c_r = \theta_{sr}\,\mathrm{d}\ln w_{sr} + (1-\theta_{sr})\,\mathrm{d}\ln w_{ur} - \lambda_{ur}\left(\kappa_r + \theta_{sr}\gamma_r\right)\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) \tag{A.14}$$

Equation (15) is also part of the aggregated version of the model (it corresponds to the upper level of the nested CES in Panel B of Figure 3), and we have $\mathrm{d}\ln P_r = \frac{1}{\varepsilon}(\,\mathrm{d}\ln Y_r - \,\mathrm{d}\ln Y)$. Combining this equation with (A.14) yields

$$\mathrm{d}\ln w_{sr} = \frac{1}{\varepsilon}(\,\mathrm{d}\ln Y - \,\mathrm{d}\ln Y_r) + (1-\theta_{sr})\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) + \lambda_{ur}\left(\kappa_r + \theta_{sr}\gamma_r\right)\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) \tag{A.15}$$

From the comparison of equations (A.10) with (A.13), (A.11) with (A.14), and (A.12) with (A.15), we conclude that the impact of a labor supply shock on regional output, prices, and wages is identical in the two versions of the model if

$$\kappa_r + \theta_{sr}\gamma_r = \bar{\eta}_r \tag{A.16}$$

To obtain the impact of a labor supply shock on relative wages, it is useful to consider the unit cost function dual to (30), which is given by

$$c_r = \frac{1}{A_r}\left[b_{sr}^{\omega_r}\left(\frac{w_{sr}}{B_r}\right)^{1-\omega_r} + b_{ur}^{\omega_r}w_{ur}^{1-\omega_r}\right]^{\frac{1}{1-\omega_r}} \tag{A.17}$$

where A_r and B_r are defined in the main text. Shephard's lemma yields conditional labor demand functions:

$$L_{sr} = \frac{1}{A_rB_r}\left(\frac{w_{sr}}{A_rB_rb_{sr}c_r}\right)^{-\omega_r}Y_r, \qquad L_{ur} = \frac{1}{A_r}\left(\frac{w_{sr}}{A_rb_{ur}c_r}\right)^{-\omega_r}Y_r$$

Relative labor demand is given by

$$\frac{L_{sr}}{L_{ur}} = \frac{1}{B_r}\left(\frac{w_{sr}}{w_{ur}}\frac{b_{ur}}{B_r b_{sr}}\right)^{-\omega_r}$$

Setting relative labor demand equal to relative labor supply and differentiating the resulting equation yields

$$\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right) = (\omega_r - 1)\,\mathrm{d}\ln B_r - \omega_r\,\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right), \qquad \mathrm{d}\ln B_r = \lambda_{ur}\gamma_r\,\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$$

This enables us to provide the equivalent of Proposition 1" in the aggregate representation of the model:

$$\mathrm{d}\ln\left(\frac{w_{sr}}{w_{ur}}\right) = -\left(\frac{1-(\omega_r - 1)\lambda_{ur}\gamma_r}{\omega_r}\right)\mathrm{d}\ln\left(\frac{\bar{L}_{sr}}{\bar{L}_{ur}}\right)$$

This expression (which captures the impact of a change in relative labor supply on relative wages in the aggregate version of the model) is identical to the expression in Proposition 1" if $\omega_r = \sigma_r$ and $(\omega_r - 1)\lambda_{ur}\gamma_r = (\varepsilon - 1)\lambda_{ur}\Phi_r$, or if

$$\omega_r = \sigma_r \quad \text{and} \quad \gamma_r = \Phi_r\left(\frac{\varepsilon - 1}{\sigma_r - 1}\right) \tag{A.18}$$

Finally, combining equations (A.16) and (A.18) yields the following expression for κ_r, in terms of the structural parameters:

$$\kappa_r = \bar{\eta}_r - \theta_{sr}\Phi_r\left(\frac{\varepsilon - 1}{\sigma_r - 1}\right) \tag{A.19}$$

A.8 Quantification Exercise: Figures

Figure A.1: Degree of Heterogeneity Between Firms in Region r (Ψ_r) in 2004

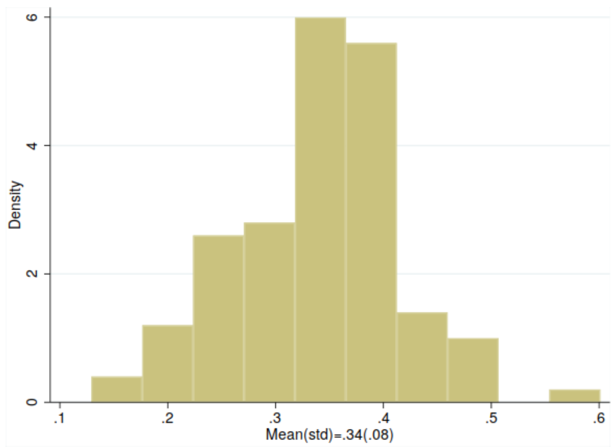
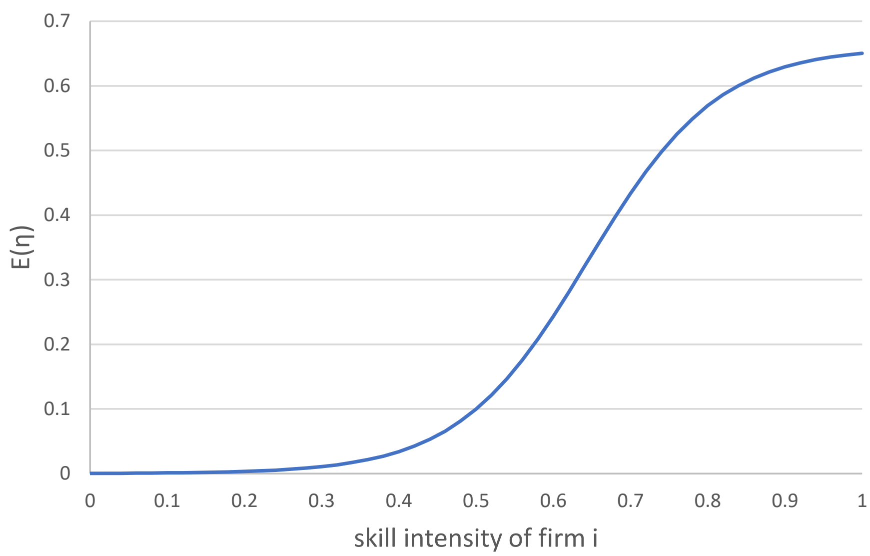


Figure A.2: The Relationship Between s_{si} and η_i



A.9 Sensitivity Analysis of the Instrument

In this subsection, we test the validity of our instrument following Borusyak et al. (2022). Specifically, we allow the possibility that shares can be endogenous and provide evidence that the shift can reasonably be considered as exogenous. To do a proper analysis, we perform the test separately for the numerator (skilled workers) and denominator (unskilled workers) components of the instrument. First, we provide descriptive statistics of the instrument in Table A.1. We can appreciate that there is enough variation across origin countries in terms of shift, with an inverse HerfindahlHirschman Index (HHI) of more than 24, a large mean, standard deviation and interquartile range for both skilled and unskilled workers. Note that shares are the same for both categories because the education level of the immigrant was not recorded in the statistics.[32]

Table A.1: Instrument Summary Statistics Across Origin Countries

	Skilled	Unskilled
Mean	8.576	8.950
SD	1.103	1.165
IQR	1.023	1.112
1/HHI	24.812	24.812
Largest share	0.076	0.076
# countries-years	221	221
# countries	38	38

Second, we run different falsification tests at the country and regional levels, and we conduct a regional pre-trend analysis. Specifically, we regress different potential confounding factors at the country and regional levels on the shock. Panel A shows that import patterns (the only available variable that can vary across origins and years) is not likely to be a confounding factor. Panel B shows that the share of women, the average wage, and the employment size in terms of total wage bill are also not correlated with the instrument. Finally, regressing the pre-period (1994–2002) employment by education on the shift-share instrument, we find that our instrument does not predict past trends in employment at the firm level. Therefore, these tests suggest that our IV approach leverages exogenous variation for identification.

[32] Since we have only about 40 origin countries, the intraclass correlation coefficient analysis is not informative for choosing the appropriate level of clustering, so we do not perform it.

Table A.2: Shock Balance Tests

	Skilled		**Unskilled**	
	Coef.	SE	Coef.	SE
Panel A: Industry-Level Balance				
Imports	-0.097	(0.071)	-0.063	(0.037)
Panel B: Regional-Level Balance				
Share of women	-0.007	(0.004)	-0.009	(0.005)
Average wage	-0.008	(0.002)	-0.007	(0.004)
Employment size	-0.016	(0.019)	-0.011	(0.032)
Panel C: Pre-trends				
1994–2002 Employment	0.002	(0.002)	0.006	0.038