The effect of immigration on natives' incomes in Finland

Benjamin Sutton

School of Business

Thesis submitted for examination for the degree of Master of Science in Economics.

Espoo 27.5.2024

Supervisor and advisor

Prof. Matti Sarvimäki

Aalto University
School of Business

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of the master's thesis

Author Benjamin Sutton

Title The effect of immigration on natives' incomes in Finland

Degree programme Economics

Major Economics **Code of major** BIZ3240

Supervisor and advisor Prof. Matti Sarvimäki

Date 27.5.2024 **Number of pages** 47+2 **Language** English

Abstract

I study the impact of immigration on natives' earnings along their income distribution in Finland. By exploiting spatial and temporal variation in inflows of foreign-background migrants combined with a shift-share instrument I find that immigration has a negative and statistically significant effect on natives below the 40^{th} income percentile, near zero effects on those between the 40^{th} and 85^{th} percentiles, and slightly positive but statistically insignificant effects on the those in and above the 90^{th} percentile. The effects are largely more intense further away from the median. The implication is that immigrants are more substitutable with low-skill Finnish labor and more complementary to high-skilled natives which results in immigration increasing income inequality at least in the short-run.

Keywords Immigration, Natives, Incomes

Aalto University
School of Business

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Pro Gradu -työn tiivistelmä

Tekijä Benjamin Sutton

Työn nimi Maahanmuuton vaikutus kantaväestön vuosituloihin suomessa

Koulutusohjelma Taloustieteen maisteriohjelma

Pääaine Taloustiede **Pääaineen koodi** BIZ3240

Työn valvoja ja ohjaaja Prof. Matti Sarvimäki

Päivämäärä 27.5.2024 **Sivumäärä** 47+2 **Kieli** Englanti

Tiivistelmä

Tutkin maahanmuuton vaikutusta suomalaisen kantaväestön vuosituloihin eri kohdissa palkkajakaumaa. Hyödyntämällä alueellista ja ajallista variaatiota maahanmuuttajataustaisten maahanmuuttajien virroissa yhdessä 'muutos-osuus' välinemuuttujan kanssa havaitsen maahanmuuton vaikuttavan negatiivisesti ja tilastollisesti merkitsevästi kantaväestön jäseniin, jotka ovat 40:n prosentiilin alapuolella vuositulojakaumassa. Tulotasoltaan 40:n ja 85:n prosentiilin väliin sijoittuville natiiveille maahanmuuton vaikutus on lähes olematon, ja korkeamalle sijoittuville hieman positiivinen, muttei tilastolllisesti merkittävä. Maahanmuuton vaikutus on yleisesti vahvempi mitä kauempana natiivi on mediaanituloisuudesta. Yhdessä tulkittuna nämä havainnot viittaavat maahanmuuttajien olevan substituutteja matalatuloisille suomalaisille ja komplementteja korkeatuloisille suomalaisille, minkä takia maahanmuutto kasvattaa tuloeroja näiden ryhmien välillä ainakin lyhyellä aikavälillä.

Avainsanat Maahanmuutto, Kantaväestö, Vuositulot

Contents

1 Introduction

The annual number of immigrants to Finland has risen steeply since the mid 2000s with the share of immigrants in the population increasing from 2.2% in 2000 to 9.1% in 2022 (StatFin, 2023b). While often argued to be necessary in patching up labor supply in the face of low birth rates and an aging population, this rapid increase has brought immigration and the economy to the forefront of political discussion. In particular, questions about the effects of immigration on native workers receive much attention. My results contribute to discussion around these issues.

Textbook economic models describe possible effects well. Immigration increases labor supply which implies less capital per worker and lower productivity. This results in downward pressure on incomes. Once capital stocks adjust productivity returns to near original levels leaving long-run effects close to zero (G. Borjas, 2010). Furthermore, if immigrants contribute positively to technological development and productivity gains or if they push natives toward higher-skill occupations they may even have a positive effect on natives' incomes. Overall effects thus depend on the degree of substitutability between natives and immigrants, the elasticity of labor and capital supplies, and differences in native and immigrant skill compositions, but are generally expected to be modest (Dustmann et al., 2005). However, these small average effects can conceal meaningful heterogeneity. In particular, while natives for whom immigration increases competition in the labor market can expect adverse effects on income, those with skills more complementary to those of immigrants may benefit (Edo, 2019). Therefore, when investigating the effects of immigration on natives it is not enough to consider average effects but effort should be taken to identify potential winners and losers.

I do this using Finnish administrative data between the years 2001 and 2018. Following the method of Dustmann et al. (2013), I employ a shift-share instrument in a fixed effects model to estimate separately the effect of a change in the ratio of foreign-background immigrants to natives in the local labor force on the income of a native earner in the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentile of the local native income distribution. Local labor markets are defined as municipalities and travel-to-work areas (TTWAs), while

foreign-background immigrants as those without Finnish citizenship upon arrival and a mother tongue other than Finnish, a more restricted group than everyone foreign-born. Essentially, I compare changes in income levels in a location with large flows of immigrants to changes in otherwise similar locations with low immigration. The estimates on the lower percentiles will reveal how immigration affects relatively low-skilled natives, while the greater percentiles reflect effects on high-skilled natives. The instrument exploits the tendency of immigrants to move to locations with an existing population from their country of origin to generate variation in immigration decisions that is independent of local labor market conditions (Altonji & Card, 1991). Hence, it addresses the bias caused by non-random assignment of immigrants across regions. I also investigate the short and long-term dynamics of the effects of immigration using the multiple-instrumentation proposed by Jaeger et al. (2018).

My results indicate that immigration has statistically significant negative effects on natives below the 40th percentile of the income distribution, near zero effects on those between the 40th and 85th percentiles, and slightly positive but statistically insignificant effects on the those in and above the 90th percentile. More specifically, if within a year the ratio of foreign-background immigrants to natives in some TTWA were to increase from just below 2 per 100 to slightly above 8 per 100[1], the expected annual income of the 25th percentile native earner would be roughly 2200€ (constant 2018 euros) lower than without such immigration. For the 10th and 5th percentile earners these figures ares 2600€ and 1200€, respectively. The equivalent estimate for the 90th percentile earner is an increase of about 1000€, but this is statistically insignificant. Relating these observations to prior research and theoretical models leads me to conclude that the skill composition of immigrants to Finland makes them more substitutable with low-skill natives and more complementary to high-skilled natives. Thus, natives below median income are likely to be hurt by immigration while those above the median are more likely to gain or feel no effect. This can widen the income distribution and increase inequality. My findings are

[1]These figures are derived from the distribution of the immigrant-to-native ratio across TTWAs in 2018 and reflect the 10th and 90th percentiles of this distribution, respectively. See Table 1 for more information.

robust to a range of deviations in the underlying population and model specification, and I find evidence that short-run effects can be even more severe than the overall estimates indicate, particularly for natives between the 10^{th} and 25^{th} income percentiles.

I contribute to the literature in two ways. First, I combine the innovative approach of Dustmann et al. (2013), which analyses effect heterogeneity among natives without pre-assignment of immigrants into skill-cells, with a method designed to address the long-term dynamic adjustment critique of shift-share instrumentation outlined by Jaeger et al. (2018). Essentially, if an economy takes many years to adjust to the shock caused by immigrants, then a shift-share instrumentation technique may confound the effects of current-day immigrants on natives with ongoing adjustment to earlier migration. Jaeger et al. (2018) design a multiple IV procedure to address this, which I adapt to my model. My results make the case that the estimates of the conventional IV model may be lower bounds for short-term effect intensity, at least for 10^{th} percentile earners. Hence, future literature employing shift-share instrumentation ought to take dynamic adjustment into consideration. Second, due to excellent Finnish data collection practices my dataset is more detailed and thorough than in most prior work, which increases statistical power. It also enables me to specifically study immigrants of a foreign background. Rather than simply tagging anyone foreign-born as an immigrant, I require that they do not have Finnish citizenship on arrival and that they do not speak Finnish as their first language. This creates a more focused collection of immigrants that better represents the group that dominates political discussion.

The next section expands on previous literature and what immigration looks like in Finland. I follow that with a description of my data and method in sections 3 and 4, which enable the subsequent discussion of results in section 5. Section 6 presents robustness checks for the primary results. Section 7 concludes.

2 Background

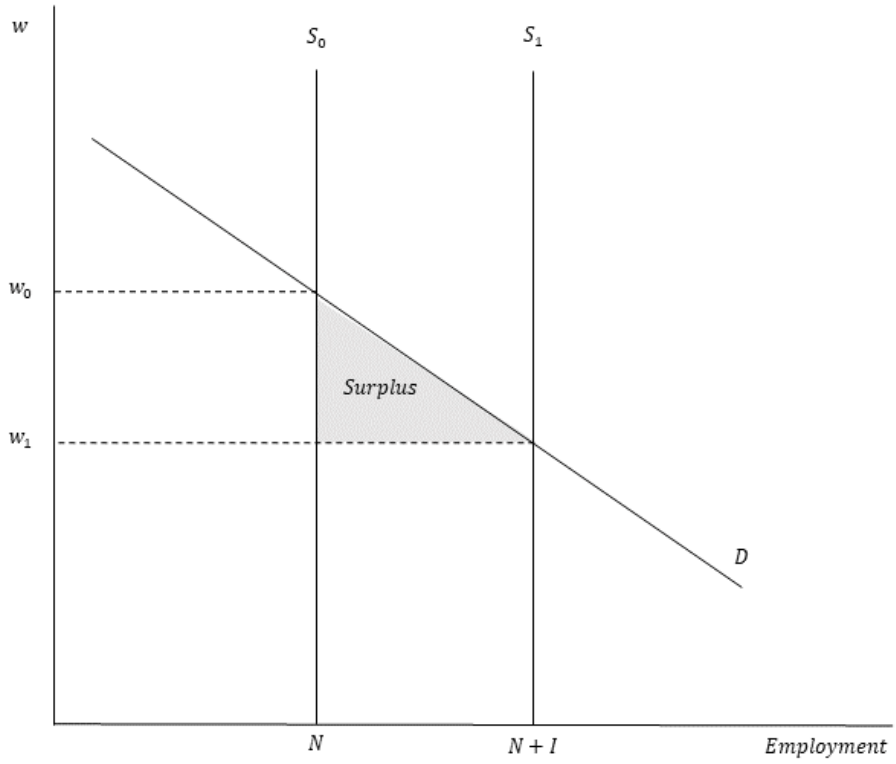
To help contextualize and explain my method and results, this section provides a brief theoretical overview of how and why immigration ought to affect natives and how it can be studied, followed by a description of immigration to Finland.

2.1 Theory

The simplest model of the labor market effects of immigration is one where an economy produces only one good using capital and labor, markets are perfectly competitive, and immigrants and natives are perfectly substitutable. In such circumstances immigration causes an increase in the labor supply thereby decreasing productivity in the short-run as there is less capital per worker. The result is lower wages and a higher return to capital. This dynamic is depicted in Figure 1, where the inelastic labor supply curve S_0 is shifted to the right by an inflow of I immigrants. In the short-run the labor demand curve D does not move causing the wage level to fall from w_0 to w_1. Importantly, the market wage is determined by the marginal product of the last immigrant worker employed. More productive immigrants will therefore be paid less than their productivity, creating whats called the immigrant surplus. This surplus, depicted as the shaded area in Figure 1, represents a net increase in national income as the gains to firms are greater than the lost wages of natives. In the long-run capital levels adjust to the increased rate of return and productivity is restored to its original level. Hence, the long term wage effects of immigration are expected to be zero (G. Borjas, 2010).

A similar story can be told via heterogeneous labor types: high and low skilled. Assuming perfectly elastic capital flows and that the skill compositions of natives and immigrants are different, immigration rather than affecting the returns to capital will affect the returns to skill. More specifically, in the skill-cells into which there is a relatively large inflow of immigrants labor supply increases and wages fall. The decreased labor cost increases profits which eventually result in wage increases for the workers in relatively unaffected skill-cells as they have now become relatively less available. In other words, natives that are more substitutable with immigrants face stiffer competition and lower wages, while

Figure 1: Wage effects of immigrants under perfect competition and homogeneous labor



Notes: the economy starts from an equilibrium where N natives are employed at wage w_0. An inflow of I immigrants shifts the inelastic labor supply curve S_0 the right. In the short-run the labor demand curve D does not move causing the wage level to fall from w_0 to w_1. The market wage is determined by the least productive immigrant that is employed creating the immigrant surplus. Source: Borjas (2010) p.184

natives who have skills complementary to immigration can expect productivity and wage increases (Edo, 2019). These distributional effects can be sustained even after adjusting for capital flows (Dustmann et al., 2016).

As differences in native and immigrant skill compositions drive the distributional effects of immigration, a brief description of the large literature on selection into migration will help understand what they are likely to be. Most of the literature is based on a Roy-model of migration which predicts that self-selection into migration is driven by the returns to skill available in various destinations. High-skilled workers will prefer a higher return to skill, essentially implying they will choose destinations that have greater inequality than their origin. The converse is true for low-skilled workers (G. J. Borjas, 1987). Hence, when studying the effects of immigrants on natives, understanding the returns to skill in the sending and receiving countries will help predict the skill composition of immigrants.

Finally, recent research has been relaxing assumptions on perfect competition and allowing firms to have market power. Specifically, Amior and Stuhler (2024) investigate the effects of immigration on the monopsony power of firms. They find that firms can make lower wage offers to migrant labor than to natives thereby reducing the employment of the latter. Hence, the employment and therefore income effects of immigration on natives can be stronger than expected under competitive models, giving further importance to empirical research quantifying them and identifying underlying dynamics.

2.2 Empirical literature

The effects of immigration on natives wages are typically identified by exploiting variation in immigrant inflows across skill-cells, regions, or both (Edo, 2019). Outcomes are determined by comparing how wages develop in a skill-cell or location with lots of immigration against one with relatively little. Assuming the units of comparison would have developed similarly in the absence of immigration, any differences in wages can be attributed to immigration. However, the skill-cell and spatial methods measure slightly different effects - the skill-cell approach only measures relative wage effects within groups, while the spatial approach captures total wage effects. Furthermore, sorting immigrants into skill-cells can

be misleading as there is evidence that immigrants tend to downgrade their occupation upon entry implying that they are not competing with natives of the same skill but in fact with ones a step below. Hence, the pure spatial approach has a slight methodological advantage (Dustmann et al., 2016) and I adopt it here. As I intend to study heterogeneity in natives' responses to immigration, some way of separating effects by skill is nevertheless required. Dustmann et al. (2013) address this by constructing separate models for different percentiles of the native wage distribution. For example, the effect of immigration on the 5^{th} and 95^{th} percentile native earner are estimated separately but with the same explanatory variable. No *ex-ante* sorting of immigrants by skill is required. Instead, if the skill composition of immigrants is different from that of natives it will be reflected by a narrowing or widening of the income distribution. Laborers in income percentiles that experience a negative effect of immigration are thus revealed to face more competition from immigration, while those enjoying a positive effect are likely to have immigrant-complementary skills. With this method, Dustmann et al. (2013) find that in the UK natives below the 20^{th} income percentile are negatively affected by immigration while those above the 40^{th} income percentile are positively affected. I contrast my findings to these.

The validity of the spatial approach faces a central issue - immigrants are not randomly allocated across regions. Instead, they tend to prefer locations in the destination country where their expected incomes are highest, creating simultaneity between labor market conditions and immigrant inflows that biases estimates of immigrant effects upward (Edo, 2019). The literature addresses this in two main ways: natural experiments and instrumental variables (IVs). The former studies sudden and unexpected migrant flows, often caused by political reasons, which can thereby be considered exogenous to local labor conditions. Examples include the Mariel Boatlift in 1980 (Card, 1990), the fall of the Berlin wall in 1989 (Glitz, 2012), and flows of refugees (Foged & Peri, 2016). The example most relevant to my work is provided by Kuosmanen and Meriläinen (2022), who study the consequences of the eastern enlargement of the EU in 2004 on native Finnish construction workers. By analysing regional and occupational variation in the density of posted workers after the enlargement they find that natives more exposed to such immigration faced 9% lower incomes relative to their less exposed counterparts. I will contrast this finding to my

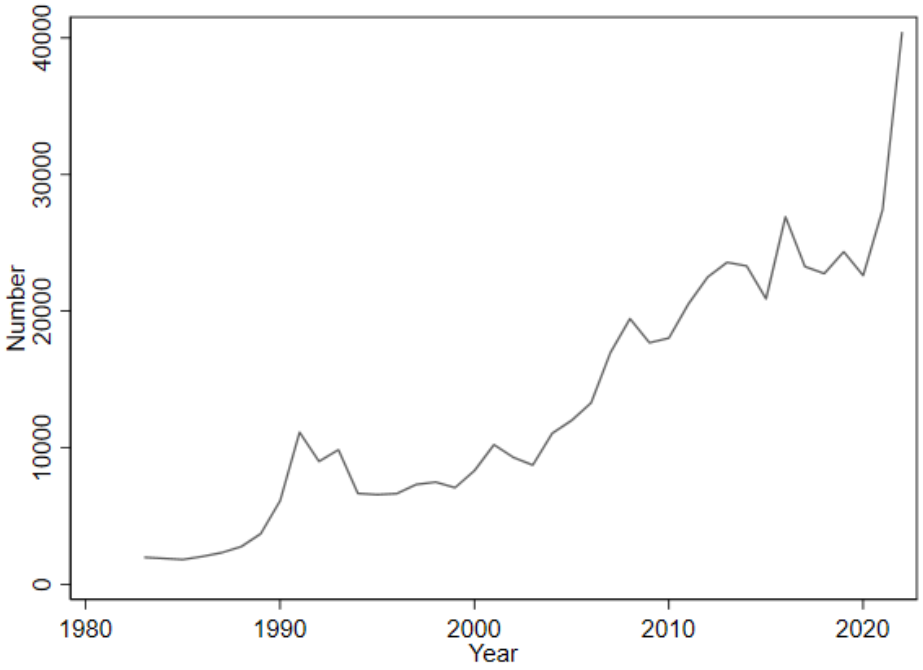
own in the results section. While natural experiments provide crisp identification, they can be hard to find and may estimate shock specific effects that differ from those of everyday immigrants. Hence, IV strategies are more common. Typically they employ a shift-share instrument which exploits the enclaving tendency of immigrants to generate variation in regional settlement that is independent of local labor markets (Altonji & Card, 1991; Card, 2009; Jaeger et al., 2018). All else held equal, immigrants tend to prefer to settle in regions with a strong existing community of people from the same country of origin (i.e. country of birth). Hence, under certain conditions, the distribution of past immigrants can be used to predict current settlement independently of regional income levels. I use a shift-share instrument in my analysis. Its motivation and flaws are investigated more thoroughly in the methodology section.

2.3 Immigration to Finland

The annual number of foreign-born immigrants to Finland is depicted in Figure 2. For most of its history Finland was a country of net emigration, and until the mid 2000s immigration to Finland was rather small in volume (Martikainen et al., 2013). Since then migrant flows have increased greatly and in 2022 immigrants made up roughly 9.1% of the population (StatFin, 2023b). These increases are largely explained by the 2004 enlargement of the European union, which reduced barriers to entry for many eastern block citizens, rising flows of refugees and other forced migrants particularly after 2015, as well as Finnish economic development making it a more attractive destination. In the face of this large increase, questions about the effects of immigration on the Finnish economy and natives have gained in importance and attention.

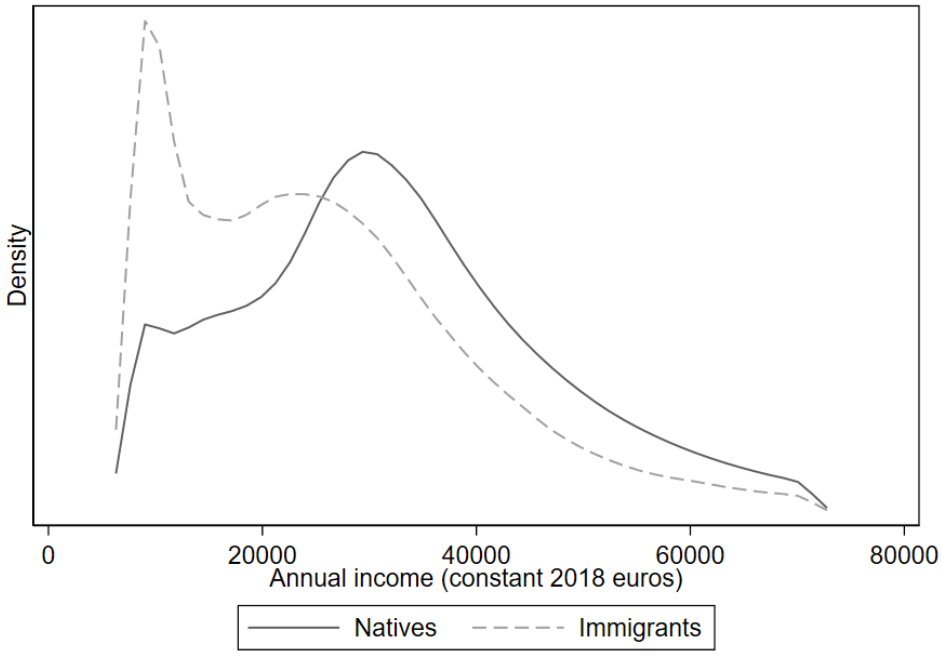
The most common countries of origin of immigrants to Finland are: Sweden, Russia, Estonia although recently inflows from Somalia, Iraq and India have increased significantly (Martikainen et al., 2013; StatFin, 2023a). Compared to these countries, excluding Sweden, Finland has relatively low-income inequality (WID, 2018). Hence the Borjas (1987) model predicts that immigrants to Finland are negatively selected and ought to be less skilled than natives. This prediction is supported empirically as the income distribution of immigrants appears to be systematically to the left of the natives', as shown in Figure

Figure 2: Annual number of foreign-background immigrants to Finland



Note: The figure shows the flow of foreign-background immigrants to Finland. An immigrant is considered foreign-background if they are foreign-born, do not have Finnish citizenship on arrival, and do not speak Finnish as their mother tongue. Source: author's calculations.

Figure 3: Annual earned income distributions of natives and immigrants



The figure represents the income distributions of natives and immigrants in the labor force between 2001 and 2018. The unit of measurement is euros. Source: author's calculations.

3. Research on educational attainment of the immigrant population indicates that the proportion of high education is similar among Finnish natives and immigrants (Sutela & Larja, 2015), but this does not necessarily imply that immigrants are represented equally in high-skilled professions. Indeed, some of this concentration in low-skill occupations may be due to barriers to entry such as language rather than strictly skill (Forsander, 2013), resulting in immigrants downgrading their occupation relative to their skill (Dustmann et al., 2013). Nevertheless, low-income Finnish natives are likely to be more substitutable with immigrant labor and hence will compete more with migrants in labor markets, while higher income natives are predicted to be more complementary to immigration. My results will test these predictions.

3 Data

I use Finnish administrative data from 2001 to 2018 as the foundation for my dataset. It contains individual level observations of every Finnish resident, and I combine basic information with employment and income data, as well as a register of migration. I focus only on the working age (15-74) population that are in the labor force (employed or unemployed). I aggregate the individual level data by municipality and year to enable spatial and temporal comparisons of the effects of immigration. I also aggregate the data by travel-to-work area (TTWA) also known as a commuting zone. A municipality is included in a TTWA if at least 10% of its labor travels daily into the central municipality of that area (StatFin, 2018). Such areas cover typical workplace commutes and can therefore be better spatial representations of an individual's labor market than their municipality of residence - it is easy to live in one municipality and work in another, but not so for TTWAs. I use the 2018 definitions of TTWAs because it is the final year of my dataset. Additionally, some municipalities that exist in earlier time periods later merge with others. To adjust to this, any municipality that eventually undergoes a merger is treated as having always been merged. Finally, while the data contains 309 municipalities in total, only 212 are covered by the 36 TTWAs from 2018. A municipality is not part of any TTWA if there is no nearby municipality to which at least 10% of its labor force commutes and if they do not receive many commuters from elsewhere. They tend to be small and rural, and face little immigration. To harmonize analysis between different location definitions, I drop all non-TTWA municipalities from the data set.

In most of the literature (e.g. Card (2009), Dustmann et al. (2013), and Foged and Peri (2016)) immigrants are defined as anyone foreign-born. This leaves some room for noise. For example, foreign-born children of expats that later return to Finland would be considered immigrants, but it is clear they would have some advantages in settlement and integration compared to immigrants with no prior relations to Finland. Such immigrants with Finnish backgrounds are likely to be more substitutable with natives than others, potentially due to linguistic abilities, and hence their effects on natives may be different than those of immigrants with entirely foreign backgrounds. As such, I use the additional

conditions that immigrants must not have Finnish citizenship in their year of arrival and do not speak Finnish as their first language. These conditions should reduce noise in the estimation. In addition to country of birth, citizenship, and language information, I also require country of birth information from immigrants, as this is necessary for constructing the instrument. Natives are defined as anyone born in Finland.

Annual real earned incomes of natives in the labor force form the response variable. Earned income is defined as the sum of salary and entrepreneurial incomes, benefits and capital gains are ignored. I adjust the income data to inflation so that they are measured in constant 2018 euros (StatFin, 2024). From this raw data, the annual income placing an individual in the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of native earners in a specific location[2] are calculated for each year. These values answer the question 'how much did the x^{th} percentile earner make in location l in the year t?' Most prior literature focuses on the impact of immigration on hourly wages (Dustmann et al., 2016), but due to features of the Finnish collective bargaining based wage-setting system (Asplund, 2007), it is likely that the income effects of immigration will be felt in reduced working hours and employment rather than wages. Indeed, Edo (2015) finds such a result in the similar wage environment of France. Annual earned income is a product of both wages and working hours, and hence ought to capture the effects of interest better than wage information. Notice that the calculated percentile specific incomes may appear smaller than expected due to the mass of unemployed and zero income natives included at the bottom of the income distribution.

The key explanatory variable is the ratio of immigrants to natives in the labor force of a location at a certain time. Its distribution is summarized in Table 1. For example, the municipality which in 2018 was in the 90^{th} percentile of municipalities with regard to labor force immigrant-to-native ratio had roughly 8 immigrants per 100 natives in the labor force. The large increases between 2001 and 2018 reflect the growing volume of immigration depicted in Figure 2. However, it is evident that areas with relatively few immigrants to begin with have not seen as much growth as those with a head start. This

[2] I use the word location throughout my text when the point is applicable to both municipalities and TTWAs.

Table 1: Percentiles of immigrants per 100 natives

	Municipality		TTWA	
Immigrants per 100 natives	2001	2018	2001	2018
mean	0.868	4.691	0.933	5.060
1st percentile	0.093	0.694	0.214	1.511
5th percentile	0.147	1.105	0.223	1.608
10th percentile	0.202	1.347	0.299	1.816
25th percentile	0.330	1.898	0.401	2.577
50th percentile	0.483	3.151	0.666	3.612
75th percentile	1.001	5.378	1.196	6.214
90th percentile	1.935	10.577	1.777	8.329
95th percentile	2.890	14.724	2.772	16.835
99th percentile	4.233	21.535	3.281	21.638

Notes: The table presents the mean and select percentile values of the immigrant-to-native ratio across locations. It shows how many immigrants per 100 natives in the labor force are required to for a location to land in a certain percentile of immigrant density.

is consistent with enclaving tendency and local labor conditions influencing location choice. The immigrant-to-native ratio is preferred to using simple inflows of migrants as a variable because it mitigates the effects of uneven distribution of immigrants across locations. An inflow of 100 migrants has very different implications depending on the size of the native labor force. Normalizing the flows by native populations produces results that are more comparable across locations. However, changes in this ratio capture not just immigrant inflows but also potential native out-migration. Fortunately, regional labor mobility in Finland is relatively modest (Poghosyan, 2018) and comparing results produced at the municipal level against the TTWA level will provide insight into this - it is more costly to move between TTWAs than municipalities meaning that large differences in estimates can be attributed to native relocation (Biavaschi et al., 2018).

The location and time specific control variables are: average age and sex ratio of the

Table 2: Means of control variables

	Natives		Immigrants	
Variable	2001	2018	2001	2018
Age	40.7	42.2	36.7	39.6
%female	48.9	49.6	44.2	44.6
%High education	17.4	32.9		
%Medium education	60.0	57.6		
%Low education	22.6	9.5		

Notes: Average age, share of women, and educational attainment shares of working age natives and immigrants in the labor force.

native and immigrant labor forces, as well as the shares of the native population with high, medium, and low educational attainment[3]. Education information is unreliable for immigrants (Sutela & Larja, 2015), hence I omit it. These variables are summarized in Table 2.

[3]High educational attainment is defined as having a bachelor-level degree or above, medium education is defined as having completed secondary education and low level as anything less.

4 Methodology

Following Dustmann et al. (2013) the effects of immigration on natives' incomes will be estimated using a first-differenced fixed effects model, presented in Equation 1.

$$\Delta I_{plt} = \alpha_t + \beta_p \Delta R_{lt} + \Delta \mathbf{X}_{plt} + \Delta \varepsilon_{plt} \tag{1}$$

The dependent variable is the change in yearly earned income I of the p^{th} percentile native earner in the labor force of a specific location l (municipality or TTWA) at a specific time t. The coefficient of interest β_p captures the percentile specific effect of a change in the ratio of immigrants to natives in the labor force R on annual earned income, averaged over municipalities and time. A vector of control variables $\boldsymbol{X}$ containing information on the age and sex composition of natives and immigrants and the educational attainment of natives is included in the model. Location specific fixed effects are eliminated by estimating the model in first-differences (subtracting the value of the previous observation from the current one) represented by Δ, while the term α consists of yearly dummies that control for the overall time trend. Error is represented by ε.

Assuming parallel trends and no spillover effects between locations, the model explains spatial and temporal differences in percentile specific incomes through differences in the volume of immigration. Locations that experience little immigration essentially serve as control groups for locations with lots of immigration. The model is estimated separately over different income percentiles to understand whether immigration affects high and low-skilled workers differently.

To overcome the issues of non-random assignment of immigrants across regions and the simultaneity between labor conditions and migration I use a shift-share instrument, defined in Equation 2 (Altonji & Card, 1991; Jaeger et al., 2018).

$$z_{lt} = \sum_o \frac{M_{olt^0}}{M_{ot^0}} \frac{\Delta M_{ot}}{N_{lt-1}} \tag{2}$$

The first term in the sum is the share of immigrants from country of origin o in location l at reference date t^0, which predates t. The national inflow of migrants from country o at time t is represented by ΔM_{ot}, and N_{mt-1} is the total population of location l in the previous period.

The instrument is founded on the well-documented enclaving tendency of immigrants - on balance immigrants tend to move to locations where a community of migrants from the same country of origin already exists (Abramitzky & Boustan, 2017; Bartel, 1989; Munshi, 2020). This is likely due to connections to residents in a certain location reducing the costs of settling there. Hence, enclaves formed prior to the reference date t^0 can be used to predict the settlement of later migrants by allocating immigrants from a certain country of origin across locations in similar proportions to how they were distributed at t^0 the reference date. This is the 'share' component of the instrument. Under this approach, variation in immigration volumes between locations is then driven by changes in the country of origin composition of total inflows - the 'shift'. As a stylized illustration of the mechanism, suppose all Swedish immigrants before t^0 moved to Turku, all Estonian immigrants to Helsinki, and everyone else to Tampere. Then if in some year the majority of immigrant to Finland were Swedish, the instrumental variable procedure would predict Turku to have the largest relative increase in immigration. Suppose in the next year the the country of origin composition changes so that Estonians now make up the majority. This would result in Helsinki experiencing the most immigration. In this way national level variation in the country of origin composition of immigrants generates local variation in immigrant shares. To the extent that national level immigration is independent of specific local labor conditions, changes in local immigration are exogenous to native income levels. Furthermore, when constructing the instrument I use data on total populations of natives and immigrants, not just the labor force. This ought to provide further independence from local labor conditions.

Equations 3 and 4 present the two stage least squares procedure I use in estimating the IV models, where γ_t are time fixed effects in the first stage and u_{lt} is the first stage error term.

$$\Delta \hat{R}_{lt} = \gamma_t + z_{lt} + \Delta \mathbf{X}_{lt} + \Delta u_{lt} \tag{3}$$

$$\Delta I_{plt} = \alpha_t + \beta_p \Delta \hat{R}_{lt} + \Delta \mathbf{X}_{plt} + \Delta \varepsilon_{plt} \tag{4}$$

The validity of the shift-share instrument, however, is called to question by serial correlation in local labor demand and slow adjustment of labor markets to shocks (Jaeger et al., 2018). In the case of the former, if labor market demand is strongly correlated between the reference period t^0 and the present, then the same factors that originally drew in immigrants will also attract immigrants today. This creates a link between the instrument and present labor market conditions which violates the exclusion restriction. The problem is addressed by ensuring the reference date is sufficiently lagged (Dustmann et al., 2013; Orrenius & Zavodny, 2007). Hence, I conduct my analysis so that the reference date precedes the data by at least eight years. I also test my IV models for serial correlation in residuals and fail to reject the null hypothesis of no second order correlation for most variables, which implies serial correlation is not a problem after instrumenting. The results of these tests are presented in the appendix in Table A1. As for slow adjustment, if the shock that past immigrants caused on a local economy is still being absorbed today, then the short-term effects of current migrants can be confounded by long-term effects of past ones. This is particularly true if the country of origin composition of migrants is relatively stable over time as it implies there is not much variation in the spatial distribution of new migrants over time (Jaeger et al., 2018). This causes point estimates to biased toward zero as, according to theory, short and long-term effects have opposite directions.

I investigate the relevance of this critique in my model by following a 'multiple-instrumentation' procedure, designed by Jaeger et al. (2018) to correct for slow dynamic adjustment. In this technique, the basic model is modified to include lagged values of the immigrant-to-native-ratio intended to capture long-term effects of migration. Each of these lagged terms is then separately instrumented for using equivalent lags of the instrument to address the same endogeneity issues that plague present immigrant flows. The lagged terms should thus control for any ongoing adjustment to older shocks, isolating short-run effects to the contemporary term. Specifically, I use the tenth lag of the immigrant-to-native ratio as a control for long-run adjustment, the same gap as employed by Jaeger et al. (2018). Shorter or longer lags can be used depending on how quickly an economy adjusts to shocks. The two first stages and the reduced form are represented in Equations 5, 6 and 7, where η_p is the percentile specific long term effect of immigration on incomes.

$$\Delta \hat{R}_{lt} = \gamma_t + z_{lt} + z_{lt-10} + \Delta \mathbf{X}_{lt} + \Delta u_{lt} \tag{5}$$

$$\Delta \hat{R}_{lt-10} = \gamma_t + z_{lt} + z_{lt-10} + \Delta \mathbf{X}_{lt} + \Delta u_{lt} \tag{6}$$

$$\Delta I_{plt} = \alpha_t + \beta_p \Delta \hat{R}_{lt} + \eta_p \Delta \hat{R}_{lt-10} + \Delta \mathbf{X}_{plt} + \Delta \varepsilon_{plt} \tag{7}$$

This method is taxing on the data and requires a very strong instrument. As such, through a comparison of first stages I choose to use the year 1993 as the reference date. It is the earliest reference date under which the instrument and its lags have sufficient predictive power over the immigrant-to-native ratio in both the single and multiple-instrumentation processes. The sensitivity of my results to other reference years is evaluated in section 6. The relevant location clustered first stage F-statistics are reported alongside regression results in the following section. Following Jaeger et al. (2018) I also utilize the Kleibergen-Paap rk LM test for underidentification as a further check of first stage validity. The test evaluates whether the matrix formed from the coefficient vectors of the first stage regressions is rank deficient against the alternative that it is of full rank (Kleibergen & Paap, 2006). Hence, the null hypothesis of the test is underidentification of the first stage and a failure to reject it jeopardizes the instrumental procedure.

As 1993 is the instrument reference year and I use a minimum lag of 8 years between the reference date and an observation, the final dataset spans the years 2001 - 2018.

5 Results

In this section I preset and discuss my empirical findings. The core empirical results from the single-IV procedure are considered in the first subsection. The second subsection focuses on multiple-IV and the slow dynamic adjustment critique.

5.1 OLS and IV results

Table 3 summarizes my OLS and IV regression results. To make interpretation more intuitive, Table 4 presents the same point estimates scaled to represent the change in income caused by a location with the 10^{th} percentile immigrant-to-native ratio in 2018 experiencing an exogenous inflow that brings them to the 90^{th} percentile (see Table 1). For example, the municipal IV point estimate for the 10^{th} native income percentile is -33912 which implies that, all else held constant, if a municipality with 1.347 immigrants per 100 natives (10^{th} percentile in 2018) faced immigration bringing this number to 10.577 (90^{th} percentile in 2018), the expected annual income of the 10^{th} percentile earner would be 3130€ smaller than without the immigration. The sample mean income for each percentile is also included in Table 4 to enable comparisons of relative effect strengths.

Both the first stage F-statistics and the Kleibergen-Paap tests indicate that the shift-share instrument has sufficient explanatory power over immigration both at the municipal and TTWA level. Hence, assuming the exclusion restriction holds, the IV procedure should be valid. It is noteworthy that the instrument appears to be stronger at the municipal level, as the test statistics are greater than their TTWA equivalents. This could be a result of there being less observations at the TTWA level or it can indicate that the immigrant enclaves that underpin the instrument are more distinct between municipalities than TTWAs and hence the instrument has more explanatory power of migrant flows at the smaller level. Either way, regardless of location definition the IV procedure is sufficiently strong to proceed with the analysis.

It is useful to start by comparing the OLS and IV estimates, as this helps contextualize the chosen method. The OLS estimates are an order of magnitude different from their IV counterparts, are generally closer to zero, and often have the opposite sign. This indicates

Table 3: The effect of immigration on native incomes

	Municipal		Travel-to-work area	
	OLS	IV	OLS	IV
Mean income	-1454	-14316	-2574	-8001
	(3577)	(6415)	(2649)	(7132)
5^{th} percentile	316	-19039	-3079	-17775
	(5111)	(8703)	(3965)	(5881)
10^{th} percentile	-298	-33912	-2684	-39672
	(5559)	(6882)	(4590)	(9066)
25^{th} percentile	828	-30268	-4375	-33079
	(3770)	(6521)	(4090)	(9914)
50^{th} percentile	3705	-18843	96	-8175
	(3204)	(5816)	(1988)	(5931)
75^{th} percentile	128	-14209	-1144	4393
	(6171)	(9157)	(4281)	(8992)
90^{th} percentile	-12802	-12767	-2880	15558
	(11278)	(19542)	(7318)	(14024)
95^{th} percentile	-14276	4719	-14711	26051
	(10749)	(21520)	(9929)	(22067)
Year dummies	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	3597	3597	612	612
F of excluded instruments	-	35	-	26
Kleibergen-Paap χ^2 statistic	-	6.17	-	4.79
		[0.000]		[0.029]

Notes: The models are estimated with first differences to eliminate location specific fixed effects. Year dummy variables are included in each specification to control for time fixed effects. The control variables are the average age and sex ratio of the native and immigrant populations in the labor force, as well as the educational attainment of the native labor force. The first stage F-statistics are calculated with location clustered standard errors. Location clustered standard errors in parentheses. KP statistic p-values in square brackets.

the simultaneity bias discussed earlier - immigrants do indeed factor in the labor market conditions of local areas when making their migration decisions. This is most evident in the results for incomes below the 50^{th} percentile, where immigrants are expected to be the most concentrated. There, the OLS estimates are significantly greater than the IV ones indicating that immigrants choose locations where low-skill incomes are rising, thus creating a positive relationship between the immigrant-to-native ratio and native earnings that biases these estimates upward (Biavaschi et al., 2018). The existence of this bias leads me to focus primarily on the IV results for the rest of the analysis, as they represent only the effect of immigration on the labor market and not the opposite.

A comparison of the IV results at the municipal and TTWA level is similarly insightful. Below the 50^{th} percentile the estimates are similar but above the median the TTWA coefficients are systematically larger and turn positive earlier. This is a somewhat surprising result. As discussed earlier, had all of the TTWA coefficients been larger in absolute value that their municipal counterparts that would have indicated that at the municipal level some of the effects of immigration were being absorbed by increased native out-migration rather than wages. But the similarity of the estimated effects on low-income between the two location indicates that such out-migration is probably not occurring in large volumes. Indeed if any group were to be incentivised to move out under immigration it would be the lowest income natives. Nevertheless, due to the underlying logic of TTWAs being better representations of labor markets than municipalities, TTWA results will be the primary focus of my analysis. Municipal results will also be addressed whenever necessary and comparing the two can serve as a robustness check while moving forward.

With these observations in mind, the results can be related to the central question of this paper - how does immigration affect natives of different income levels? The percentile specific IV estimates indicate that immigration has a negative and statistically significant effect on the earnings of natives in the 5^{th}, 10^{th}, and 25^{th} percentiles of the native income distribution. For those or above the median the effects are zero or slightly positive and statistically insignificant. Drawing on theory this serves as evidence that foreign-background immigrants to Finland are more substitutable with low-income natives and more complementary to higher income natives. An influx of migrants therefore represents

increased competition in the local labor market for low-income natives, which places downward pressure on their average real earnings relative to unaffected locations. These negative effects do not imply that low-income natives are necessarily made poorer by immigration on average, but rather that their incomes do not increase by as much as they would have in the absence of it. Importantly, I cannot say how exactly these negative effects manifest, whether it is reduced wages or working hours, but simply observe the total effect.

Somewhat surprisingly, the estimates imply that high-income natives do not as unambiguously benefit from the immigrant surplus as expected. At the median and above the estimated coefficients are largely statistically insignificant at the 5% level. This may indicate relatively large heterogeneity in how high-income natives experience immigration within a location which brings the average effect closer to zero. I identify three potential sources of such differences. First, it is possible that the gains of immigration do not flow to all higher income natives as smoothly as imagined and are instead concentrated in specific industries or at even higher earnings levels. Second, while most immigrants are lower-skilled than natives, high-skill immigration does occur, particularly in tech industries. Hence high-income natives in locations with relatively large amounts of high-skilled immigrants may also face increased competition and downward pressure on income. Indeed, if a few high-income natives become unemployed due to immigration, then even if everyone else experiences a modest gain, the average effect can be close to zero. Third, as in the simple immigrant surplus model presented in the beginning of Section 2, the benefits of immigration could be accrued through greater return to capital, which I do not observe. However, it is also possible that I do not detect positive effects of immigration on high-income natives because they spill over across locations. This would happen if high-skilled laborers are able to provide their services to a greater area than low-skilled labor and could therefore benefit from immigration that occurs relatively far away.[4] Under such

[4]As a stylised example consider construction workers and architects in two locations A and B that are far away from each other. Suppose location A experiences a large inflow of low-skilled immigrants. This would increase competition for construction jobs in A but likely would not have an effect on native builders in B as the spatial mobility of construction workers is limited to few hour commutes. Thus, comparing the development of construction worker incomes between A and B would reveal the effect of immigration. An

circumstances high-income natives in locations with low volumes of immigration would not be good counterfactuals for high-income natives in locations with high immigration and comparisons between them would not reveal the positive effects of immigration. The aforementioned differences between municipal and TTWA results can be interpreted in this light. Spillovers are more likely between municipalities than TTWAs resulting in the municipal coefficients detecting less positive effects than TTWA ones and therefore being smaller than their TTWA counterparts at and above the median. This could also explain why the estimated effects on the mean native are further away from zero than expected - positive effects may be partially lost in inter-location comparisons.

In terms of magnitude, the absolute values of the IV estimates indicate that the effects of immigration are more intense further away from the median. For example, as is presented in Table 4, if a TTWA with 1.816 immigrants per native in 2018 were to experience an inflow bringing them to 8.329 immigrants per native, the expected annual income of the 90^{th} percentile earner would increase by 1013€ while the 95^{th} percentile earner can expect a larger increase of 1697€. The trend holds even if the coefficients are compared to the mean incomes, also presented in Table 4. This indicates that natives at the lowest end of the distribution are the most substitutable with immigrants in terms of skill sets, while those at the highest end are the most complementary. This observation is in line with the earlier discussion of likely negative-selection among immigrants to Finland. It is also consistent with empirical findings from other countries in which immigrant and native skill composition is explicitly accounted for (Altonji & Card, 1991; Orrenius & Zavodny, 2007; Peri & Sparber, 2009). The exception to this trend is the difference between estimated coefficients at the 5^{th} and 10^{th} income percentiles - the former is smaller in absolute value than the latter. Three possible explanations are likely. First, native workers at the 5^{th} income percentile can be more likely to leave the labor force in the face of immigration, compared to workers in the tier above. As such the effect of immigration is not absorbed

architect, on the other hand, is able to provide their services across a larger area, perhaps even exceeding a TTWA. Hence, the increase in construction workers in A could result in more construction projects in A and therefor more demand for architects. Then architects in both A and B could benefit from the immigration to A, and hence a comparison of architect incomes in A and B would reveal little.

Table 4: The scaled effect of immigration on native incomes

		Municipality		Travel-to-work area	
	Mean income	OLS	IV	OLS	IV
Mean income	31632	-134	-1321	-168	-521
		(330)	(592)	(173)	(464)
5^{th} percentile	7388	29	-1757	-201	-1158
		(472)	(803)	(258)	(383)
10^{th} percentile	10506	-28	-3130	-175	-2584
		(513)	(635)	(299)	(590)
25^{th} percentile	19705	76	-2794	-285	-2154
		(348)	(602)	(266)	(646)
50^{th} percentile	29774	342	-1739	6	-532
		(296)	(537)	(129)	(386)
75^{th} percentile	39582	12	-1311	-75	286
		(570)	(845)	(279)	(586)
90^{th} percentile	52236	-1182	-1178	-188	1013
		(1041)	(1804)	(477)	(913)
95^{th} percentile	62247	-1318	436	-958	1697
		(992)	(1986)	(647)	(1437)

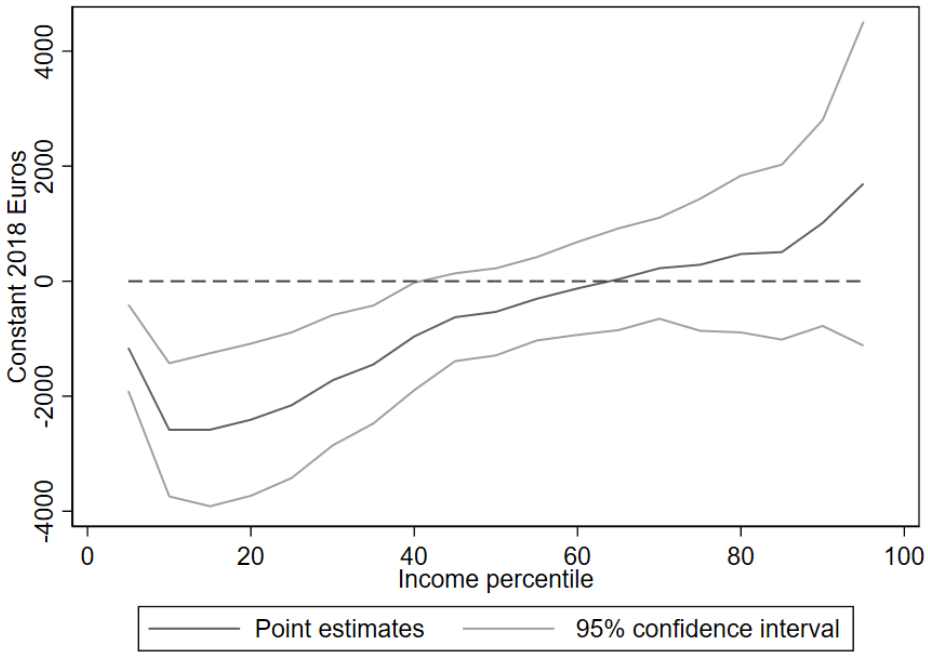
Notes: The table presents the point estimates from Table 3, scaled to represent the expected change in annual income caused by a location in the 10^{th} percentile of immigrants per native in the labor force in 2018 going to the 90^{th} percentile. The inter-TTWA and inter-temporal mean income associated with each percentile is presented to help quantify effects. Scaled standard errors in parentheses.

entirely by annual incomes, which attenuates the estimates. Second, some empirical evidence exists of immigration pushing low-skilled native workers from manual labor to more communication based occupations as linguistic and cultural differences imply that native workers have a comparative advantage in the latter (Foged & Peri, 2016; Peri & Sparber, 2009). Natives who switch jobs like this can increase their income, thus resulting in total estimated effects of immigration being closer to zero. It is possible that natives in the 5^{th} percentile are more likely to undergo such a profession upgrade than those in the 10^{th} causing the reversal of effect intensity. Third, it may simply be that 5^{th} percentile earners have less total income to lose, making the coefficient smaller in size even if relative effect intensity is greater. This point, however, is weakened by comparing the absolute value of the scaled coefficients to the mean incomes of 5^{th} and 10^{th} percentile earners. In the specified scenario the 10^{th} percentile earner is expected to lose 2584€ which is 25% of their mean income. The equivalent estimate for the 5^{th} percentile native is 1158€ which is 16% of their mean income.

My key takeaways are best summarised by Figure 4, plots the scaled TTWA IV point estimates for every fifth income percentile. The Figure confirms the trends suggested in the regression tables - immigration has negative effects on below median income natives, and zero or slightly positive effects on above median natives. Statistically significant effects are found only below the 40^{th} percentile.

Compared to the work of Dustmann et al. (2013) the sentiment of my results is similar - there is a positive relationship between a native's income level and the effect of immigration on them. The spread and statistical significance of the effects are drastically different, however, as they find negative and significant effects on natives below the 20^{th} income percentile and positive and significant effects above the 40^{th}, while I do not have statistically significant positive effects at any income percentile. This can indicate that the variance of the skill distribution of immigrants is greater in Finland than the UK, making them more substitutable for natives at higher income percentiles, which is consistent with the observations of Sutela and Larja (2015). Or its possible that the aforementioned dispersion of positive effects across locations is more prevalent in Finland due to the smaller overall population.

Figure 4: Scaled travel-to-work-area IV point estimates



Notes: The figure presents income percentile specific TTWA IV estimates of the expected change in native annual income caused by a location in the 10^{th} percentile of immigrants per native in the labor force in 2018 going to the 90^{th} percentile. The estimates are calculated for every fifth native income percentile starting at the 5^{th} percentile.

My findings are more consistent with those of Kuosmanen and Meriläinen (2022) who study the effects of immigration on Finnish construction workers. They find that the increased number of posted workers to Finland that followed the eastern enlargement of the EU in 2004 resulted in a 9% decrease in the annual income of more exposed natives relative to less exposed ones. They also observe that less exposed natives may even have benefited. My results are therefore similar in spirit as I also find that immigration creates winners and losers based off of the degree of substitutability, and given that native construction workers tend to be in the lower income percentiles the direction and magnitude of the effects are comparable.

5.2 Multiple-IV results

Overall, the results of the single-IV estimation indicate that immigration has a significant negative impact on native earners below the 40^{th} income percentile. This finding is slightly strengthened by analysis of the multiple-IV estimates, which are presented in Table 5. I make the case that the single-IV estimates for the 10^{th} percentile native earner may be biased toward zero by slow adjustment to earlier immigration and that therefore the short-term effects of immigration are more negative than thus far estimated. For other percentiles the existence of such a bias is unclear.

To form this argument, it is necessary to start by analysing the strength of the instruments in the multiple-IV specification. The first stage F-statistics, calculated using location clustered standard errors, range from 14 to 18. This implies that the shift-share instrument and its tenth lag together seem to have sufficient explanatory power over the immigrant-to-native ratio both today and ten years ago. Even stronger instruments with larger F-statistics would be desirable (Lee et al., 2022), but these values are sufficient to progress. The Kleibergen-Paap rk LM statistics are more concerning. In particular, with a p-value of 0.158, I fail to reject that the system of equations consisting of current and lagged immigration explained by the instrument and its tenth lag is underidentified at the TTWA level. This jeopardizes the multiple-IV procedure, and hence the TTWA estimates may not be informative. However, with a p-value of 0.013, underidentification can be rejected at the municipal level. Hence, the multiple-IV approach appears to be valid only at the

Table 5: Multiple-IV results

	Municipal			Travel-to-work area		
	IV	Multi-IV		IV	Multi-IV	
	R	R	L10.R	R	R	L10.R
Mean income	-14316	-26220	15441	-8001	2175	-4092
	(6415)	(22857)	(33553)	(7132)	(12379)	(14957)
5^{th} percentile	-19039	-9059	-12674	-17775	-62665	49898
	(8703)	(16099)	(25095)	(5881)	(35963)	(39044)
10^{th} percentile	-33912	-68307	32398	-39672	-77233	37076
	(6882)	(29212)	(40508)	(9066)	(32186)	(35456)
25^{th} percentile	-30268	-26203	9643	-33079	-33696	14861
	(6521)	(13305)	(47602)	(9914)	(20026)	(24173)
50^{th} percentile	-18843	-7445	-17382	-8175	-17137	14572
	(5816)	(20999)	(32899)	(5931)	(13350)	(17750)
75^{th} percentile	-14209	-46603	40854	4393	22492	-22076
	(9157)	(31656)	(44390)	(8992)	(26446)	(29604)
90^{th} percentile	-12767	-54080	34037	15558	4729	17356
	(19542)	(97048)	(121532)	(14024)	(30250)	(34820)
95^{th} percentile	4719	-149739	204156	26051	31779	-9570
	(21520)	(92415)	(151337)	(22067)	(62458)	(76013)
N	3597	3597	1477	612	612	252
F of instruments	35	14	18	26	16	15
Kleibergen-Paap χ^2	22.85	6.17		4.79	1.99	
	[0.000]	[0.013]		[0.029]	[0.158]	

Notes: The table presents the estimated coefficients for the ratio of immigrants to natives (R) and its tenth lag (L10.R). The multiple-IV estimates for the current and lagged term are estimated simultaneously. The models are estimated in first differences which eliminates location specific fixed effects. Year dummy variables are included in each specification to control for time fixed effects. Included control variables are the average age and sex ratio of the native and immigrant populations in the labor force, and the educational attainment of the native labor force. The F-statistic represents the joint explanatory power of both the current and past instrument over immigration. Location clustered standard errors in parentheses. KP statistic p-values in square brackets.

municipal level and I will focus primarily on those results in the following analysis.

The multiple-IV coefficients for the current and lagged effects of a change in the immigrant-to-native ratio (R and L10.R) are estimated simultaneously in the same model. Therefore, the estimates should be read together when making inferences. The multiple-IV coefficients for the current term are universally negative while the coefficients for the lagged term are mostly positive. This is in line with theoretical predictions, as the benefits of immigration are gained once capital levels readjust to the new labor supply, which can take longer than adjustments to wages and employment triggered by the same shock. The estimates are, however, rather noisy with standard errors far larger than their single-IV counterparts and a general lack of statistical significance. These large errors, likely a result of relatively weak instruments or a reflection of aforementioned within income-group heterogeneity in immigration experiences, imply that the multiple-IV estimates are not particularly informative on their own. Little can be said about the specific strengths of short and long-term effects based off them. But when comparing them to the range of single-IV estimates they can be insightful.

This is particularly true for estimates of the effects of immigration on the 10^{th} percentile earner, as the municipal multiple-IV coefficient for the current term (-68307) is statistically significant at the 5% level. It is also far smaller than its single-IV counterpart (-33912), indicating that short-run effects of immigration on the 10^{th} percentile earner appear to be more intense than represented by the single-IV approach. Furthermore, the multiple-IV point estimate for the the long term effect on the 10^{th} percentile is positive albeit statistically insignificant, which is consistent with the idea that the single-IV estimates are indeed biased upward by slow adjustment to earlier immigration. Essentially, it seems that the overall effect of an increase in the immigrant-to-native ratio on the 10^{th} percentile earner is negative and significant, and that in the short-run these effects can be even stronger after which they eventually slightly abated by positive long-run effects caused by capital adjustments. A similar story is told by estimated effects on mean incomes, albeit without any statistical significance - the overall effect is negative, but less so than the short-term one. Therefore, at least for the 10^{th} percentile results, it seems reasonable to conclude that the single-IV estimates confound short and long-term effects of immigration and that the

true short-run effects are likely more negative than represented by the single-IV results.

Whether this dynamic is at play for other income levels is unclear. For 6 out of the 8 estimates the short-run effect is negative while the long-run is positive. In the remaining two cases both estimates are negative, indicating a lack of positive long-run adjustment. Furthermore, some of the short-run estimates are closer to zero than their single-IV counterparts, which runs counter to the argument that the latter are biased upward. None of the estimates are statistically significant at the 5% level.

6 Robustness checks

The key observations drawn from the single-IV approach are that immigration has negative and statistically significant effects on low-income natives and zero or positive and statistically insignificant effects on high-income natives. Additionally, the intensity of these effects increases further away from the median, with the exception of a decrease in intensity below the 10^{th} percentile. This section evaluates the robustness of these results to various changes in the underlying population and instrument definition.

6.1 Robustness to population changes

In this subsection, I reproduce the IV results using data constructed with one of the four following deviations from the baseline population:

1. Working age changed to 20-59 from 15-74
2. Only men
3. Only women
4. Only salaried workers (not self-employed)

The new population definitions are applied to both natives and immigrants.

Municipal level results are presented in Table 6 and TTWA level results in Table 7. Notice in the municipal specifications the population restrictions cause some municipalities to be omitted from analysis as they no longer posses enough immigrants in certain time periods. Thus some variation in coefficient estimates can be due to the exclusion of certain municipalities rather than estimate sensitivity to underlying population characteristics. Every TTWA is represented in every specification implying that results at that level may be better suited to analyse robustness.

Overall the results of these checks increase my confidence in the core findings. In nearly all specifications at all levels the point estimates are negative for low-income natives and then rise with income level, again with the exception of estimated effects for the 5^{th} percentile being smaller in absolute value than for the 10^{th}. Curiously the models that

Table 6: Municipal level population robustness checks

	Baseline	Age	Male	Female	Salaried
Mean income	-14316	-21534	-22802	-3908	-11200
	(6415)	(5842)	(7620)	(5724)	(7227)
5^{th} percentile	-19039	-19336	6758	-41133	-11593
	(8703)	(5502)	(13068)	(13394)	(7715)
10^{th} percentile	-33912	-32851	-26728	-48580	-31476
	(6882)	(7182)	(8726)	(12982)	(7035)
25^{th} percentile	-30268	-35965	-33951	-35843	-24388
	(6521)	(7258)	(11490)	(7362)	(12858)
50^{th} percentile	-18843	-20863	-42973	1521	-16346
	(5816)	(7726)	(14653)	(7700)	(9625)
75^{th} percentile	-14209	-20722	-37746	12158	-14018
	(9157)	(13090)	(15290)	(12778)	(13690)
90^{th} percentile	-12767	-4919	-30194	41645	7031
	(19542)	(12666)	(21807)	(17937)	(10218)
95^{th} percentile	4719	646	-14364	30743	19225
	(21520)	(15713)	(37573)	(26098)	(39844)
Year dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	3597	3570	3213	3451	3536
F of instruments	35	37	34	31	29
Kleibergen-Paap χ^2	23	23	21	21	22
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: The name at the top identifies what aspect of the population has been changed to test robustness. 'Baseline' refers to the single-IV results presented earlier. In the 'Age' model the age range of the population is decreased from 15-74 to 20-59. Columns 'Male' and 'Female' consider men and women separately. Model 'Salaried' excludes self-employed workers. Other specification details are the same as in the baseline IV model, except the controls for sex ratio are omitted in the 'Male' and 'Female' models as they are constant by construction. The first stage F-statistics are calculated with municipality clustered standard errors. Municipality clustered standard errors in parentheses. KP statistic p-values in square brackets.

Table 7: TTWA level population robustness checks

	Baseline	Age	Male	Female	Salaried
Mean income	-8001	-12130	-11495	-3337	-7614
	(7132)	(7274)	(8400)	(6773)	(7581)
5^{th} percentile	-17775	-28683	-15705	-25024	-16969
	(5881)	(7853)	(7588)	(6719)	(7145)
10^{th} percentile	-39672	-44618	-41694	-44580	-42763
	(9066)	(12337)	(9818)	(13311)	(11738)
25^{th} percentile	-33079	-33835	-32580	-39878	-33002
	(9914)	(11066)	(11462)	(15645)	(10787)
50^{th} percentile	-8175	-9345	10072	-4364	-8037
	(5931)	(5271)	(6850)	(6906)	(5763)
75^{th} percentile	4393	938	-6474	21933	4203
	(8992)	(7865)	(10719)	(8849)	(9462)
90^{th} percentile	15558	8245	1838	25563	16907
	(14024)	(13867)	(13653)	(18266)	(15367)
95^{th} percentile	26051	20871	20740	53344	25895
	(22067)	(21648)	(20104)	(28788)	(23837)
Year dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	612	612	612	612	612
F of instruments	26	22	29	21	21
Kleibergen-Paap χ^2	4.79	4.79	4.86	4.79	4.75
	[0.029]	[0.029]	[0.028]	[0.029]	[0.029]

Notes: The name at the top identifies what aspect of the population has been changed to test robustness. 'Baseline' refers to the single-IV results presented earlier. In the 'Age' model the age range of the population is decreased from 15-74 to 20-59. Columns 'Male' and 'Female' consider men and women separately. Model 'Salaried' excludes self-employed workers. Other specification details are the same, except the controls for sex ratio are omitted in the 'Male' and 'Female' models. The first stage F-statistics are calculated with TTWA clustered standard errors. TTWA clustered standard errors in parentheses. KP statistic p-values in square brackets.

consider men only are exceptions to this. The municipal 5th percentile coefficient and the TTWA 50th percentile coefficient are positive, contrary to expectations. The former could imply that men are more likely than women to be pushed by immigration into higher paying communication based occupations, resulting in more positive experiences of immigration. The latter is more vexing as the coefficients above and below it are negative. I am unable to provide any explanation why the effects of immigration would be positive on median male earners and negative on 75th percentile men, other than the result being a statistical aberration. Neither coefficient is statistically significant, supporting this view. Furthermore, at and above the median the estimates from the women only model are systematically larger than their male counterparts. Hence, it would appear that the effects of immigration on labor market outcomes can be different between men and women.

A more certain result is that immigration has negative and statistically significant effects on earners between the 5th, 10th, and 25th percentiles of the native income distribution. The estimated coefficients for these percentiles are remarkably across models particularly at the TTWA level and are statistically significant at the 5% level in all but two cases. As these estimates are so insensitive to changes in the underlying population, it is reasonable to infer that the scaled effect strengths displayed in Figure 4 could be very close to the truth at and below the 25th percentile. The 5th percentile native earner in TTWA moving from the 10th percentile of immigrant density in 2018 to the 90th percentile, can expect their annual income to be about 1200€ less than otherwise. The equivalent figures for the 10th and 25th percentile native earners are 2600€ and 2200€, respectively.

6.2 Robustness to instrument reference date

In this subsection I evaluate the sensitivity of the single-IV estimates to changes in the instrument's reference date (see Equation 2) by reproducing the models with instruments constructed with the reference year ranging from 1991 to 1995 (1993 is the baseline). As I keep constant the start and end dates of the dataset (2001-2018) changing the reference year also alters the gap between the reference and the first observation. Municipal level results are presented in Table 8, TTWA level results in Table 9.

Table 8: Municipal level reference year robustness checks

	1991	1992	1993	1994	1995
Mean income	-17874	-17321	-14316	-18603	-17687
	(5755)	(6610)	(6415)	(7264)	(6399)
5^{th} percentile	-29553	-28854	-19039	-28731	-30029
	(11874)	(15121)	(8703)	(17154)	(14767)
10^{th} percentile	-42929	-40508	-33912	-43334	-43768
	(7944)	(11539)	(6882)	(13936)	(11875)
25^{th} percentile	-39620	-35644	-30268	-35943	-38124
	(7282)	(8502)	(6521)	(8711)	(8610)
50^{th} percentile	-16789	-20071	-18843	-21855	-19358
	(6774)	(6860)	(5816)	(8318)	(6799)
75^{th} percentile	-9474	-18899	-14209	-23121	-17687
	(10979)	(10101)	(9157)	(13055)	(10081)
90^{th} percentile	6423	-17156	-12767	-25695	-16305
	(11350)	(14898)	(19542)	(21959)	(21027)
95^{th} percentile	-92	-6681	4719	-198	3983
	(17616)	(19173)	(21520)	(23982)	(21939)
Year dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	3597	3597	3597	3597	3597
F of instruments	34	27	35	24	32
Kleibergen-Paap χ^2	22	22	23	23	23
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: The year at the top identifies the reference year used to construct the instrument. The 1993 estimates are the baseline. Other specification details and the data are the same as in the baseline IV model. The first stage F-statistics are calculated with municipality clustered standard errors. Municipality clustered standard errors in parentheses. KP statistic p-values in square brackets.

Table 9: TTWA level reference year robustness checks

	1991	1992	1993	1994	1995
Mean income	-6999	-8376	-8001	-8812	-9087
	(7596)	(5908)	(7132)	(7347)	(8084)
5^{th} percentile	-18970	-19203	-17775	-16934	-17506
	(6837)	(7016)	(5881)	(5470)	(5260)
10^{th} percentile	-31056	-36639	-39672	-38941	-42307
	(11198)	(10998)	(9066)	(9852)	(10924)
25^{th} percentile	-27797	-27437	-33079	-33763	-35688
	(11969)	(9486)	(9914)	(10692)	(11736)
50^{th} percentile	-5713	-9781	-8175	-8761	-8077
	(5379)	(5946)	(5931)	(6426)	(6140)
75^{th} percentile	5226	2408	4393	3991	4497
	(8652)	(8772)	(8992)	(9214)	(9377)
90^{th} percentile	15234	10401	15558	14272	16308
	(14747)	(16013)	(14024)	(14514)	(14370)
95^{th} percentile	16152	19578	26051	18297	24766
	(26738)	(25413)	(22067)	(22807)	(22039)
Year dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	612	612	612	612	612
F of instruments	22	17	26	21	20
Kleibergen-Paap χ^2	5.15	4.67	4.79	4.54	4.95
	[0.023]	[0.031]	[0.029]	[0.033]	[0.026]

Notes: The year at the top identifies the reference year used to construct the instrument. The 1993 estimates are the baseline. Other specification details and the data are the same as in the baseline IV model. The first stage F-statistics are calculated with TTWA clustered standard errors. TTWA clustered standard errors in parentheses. KP statistic p-values in square brackets.

Overall, the estimates are rather insensitive to changing the instrument's reference date. The trend of negative and statistically significant effects for low-income natives, near zero effects on median natives, and positive but statistically insignificant effects on high-income natives holds across all specifications. Effect intensity increases with distance from the median in all the models. Furthermore, all point estimates at the TTWA level are very similar to each other. At the municipal level this is the case for the 5^{th}, 10^{th} and 25^{th} percentiles. Therefore, it appears my findings are not specific to the choice of instrument reference year or gap to the start of the data.

7 Conclusion

I investigate the effects of foreign-background immigration on the incomes of native Finns. With individual level register data from 2001 to 2018 I use a fixed-effects model to compare changes in income levels between municipalities and travel-to-work areas that experience varying flows of immigrants over time. I employ a shift-share instrument to enable causal identification. Furthermore, I utilize the multiple-IV procedure designed by Jaeger et al. (2018) to address bias in estimates caused by slow long-term adjustment to past immigration. Finally, I evaluate the robustness of my key results by analysing the stability of my estimates across various underlying populations and instrument reference dates.

I find evidence that native earners below the 40^{th} income percentile are negatively affected by immigration, natives between the 40^{th} and 85^{th} percentiles are largely unaffected, and those above the 90^{th} percentile are affected positively but statistically insignificantly. The impacts are generally more intense further away from the median and, as indicated by the multiple-IV results, may be even stronger in the short-term for low-income natives. These findings are insensitive to changes in the underlying data or instrument reference date. Overall, my results indicate that the skill-distribution of foreign-background immigrants to Finland is more complementary to high-skilled natives and more substitutable with that of lower-skilled natives. Hence, immigration increases income inequality among natives, at least in the short-run.

I contribute to the literature by further elaborating on the heterogeneity of the effects of immigration on natives of different skill levels. Any future work on the effects of immigrants on the incomes of natives must separate the experiences of low and high-skilled natives from each other. Indeed, given that long-term net effects tend to be small, understanding the distributional consequences of immigration may be more relevant to policymakers. Furthermore, by augmenting the income percentile specific modelling of Dustmann et al. (2013) with the multiple-IV technique of Jaeger et al. (2018) I provide evidence that my conventional IV estimates may be a lower bound in absolute value for short-term effects. Natives in the 10^{th} percentile of the income distribution in particular appear to face far stronger short-term effects once slow adjustment to past immigration is controlled for. As

such, further research on the topic ought to pay heed to dynamic elements of immigration and interpret results accordingly.

In light of my findings and the recent large increases in immigration to Finland, there may be reason to support low-income natives affected by immigration in the short-term to avoid significant damage. Prior research has shown that inflows of low-skilled immigrants can push natives out of manual labor toward occupations requiring more communication skills, thereby increasing their incomes (Foged & Peri, 2016; Peri & Sparber, 2009). Policymakers may wish to take steps that facilitate such transitions to maximize the potential gains from immigration while minimising growing pains.

Three avenues for future research arise naturally from my work. First, in studying total income effects I do not separate whether immigration is felt mostly in employment or wages. Given the prevalence of collective bargaining in Finnish wage setting, an analysis of which margin is most affected could be fruitful. Second, studies with access to *ex-ante* information on immigrants could help illuminate why immigrants are more concentrated in low-skilled jobs and elaborate on whether immigrants displace natives through downgrading their occupation, receiving lower wage offers or for some other reason. Third, the robustness checks indicate there may be gender differences in the effects of immigration on incomes. Identifying whether the sex composition of natives or immigrants is driving these or if they are a result of gender being associated with occupational choice may be insightful.

References

Abramitzky, R., & Boustan, L. (2017). Immigration in american economic history. *Journal of economic literature*, *55*(4), 1311–1345.

Altonji, J. G., & Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *The new immigrant in the american economy* (pp. 201–234). University of Chicago Press.

Amior, M., & Stuhler, J. (2024). Immigration, monopsony and the distribution of firm pay.

Asplund, R. (2007). *Finland: Decentralisation tendencies within a collective wage bargaining system* (tech. rep.). ETLA Discussion Papers.

Bartel, A. P. (1989). Where do the new us immigrants live? *Journal of labor economics*, *7*(4), 371–391.

Biavaschi, C., Facchini, G., Mayda, A. M., & Mendola, M. (2018). South–south migration and the labor market: Evidence from south africa. *Journal of Economic Geography*, *18*(4), 823–853.

Borjas, G. (2010). *Labor economics*. McGraw-Hill Companies,Incorporated. https://books.google.fi/books?id=zFMoAQAAIAAJ

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *The American Economic Review*, *77*(4), 531–553.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Ilr Review*, *43*(2), 245–257.

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

Dustmann, C., Fabbri, F., & Preston, I. (2005). The impact of immigration on the british labour market. *The Economic Journal*, *115*(507), F324–F341.

Dustmann, C., Frattini, T., & Preston, I. P. (2013). The effect of immigration along the distribution of wages. *Review of Economic Studies*, *80*(1), 145–173.

Dustmann, C., Schönberg, U., & Stuhler, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, *30*(4), 31–56.

Edo, A. (2015). The impact of immigration on native wages and employment. *The BE Journal of Economic Analysis & Policy*, *15*(3), 1151–1196.

Edo, A. (2019). The impact of immigration on the labor market. *Journal of Economic Surveys*, *33*(3), 922–948.

Foged, M., & Peri, G. (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, *8*(2), 1–34.

Forsander, A. (2013). Maahanmuuttajien sijoittuminen työelämään. In *Muuttajat: Kansainvälinen muuttoliike ja yhteiskunta* (pp. 220–239). Gaudeamus Helsinki.

Glitz, A. (2012). The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in germany. *Journal of Labor Economics*, *30*(1), 175–213.

Jaeger, D. A., Ruist, J., & Stuhler, J. (2018). *Shift-share instruments and the impact of immigration* (tech. rep.). National Bureau of Economic Research.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126.

Kuosmanen, I., & Meriläinen, J. (2022). Labor market effects of open borders: Evidence from the finnish construction sector after eu enlargement. *Journal of Human Resources*.

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. (2022). Valid t-ratio inference for iv. *American Economic Review*, *112*(10), 3260–3290.

Martikainen, T., Saari, M., & Jouni, K. (2013). Kansainväliset muuttoliikkeet ja suomi. In *Muuttajat: Kansainvälinen muuttoliike ja yhteiskunta* (pp. 23–38). Gaudeamus Helsinki.

Munshi, K. (2020). Social networks and migration. *Annual Review of Economics*, *12*, 503–524.

Orrenius, P. M., & Zavodny, M. (2007). Does immigration affect wages? a look at occupation-level evidence. *Labour Economics*, *14*(5), 757–773.

Peri, G., & Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, *1*(3), 135–169.

Poghosyan, M. T. (2018). *Regional labor mobility in finland*. International Monetary Fund.

StatFin. (2018). Travel-to-work areas. https://www.stat.fi/fi/luokitukset/tyossakayntial/tyossakayntial_1_20180101/

StatFin. (2023a). Official statistics of finland: Migration [ISSN = 1797-6782]. https://www.stat.fi/tilasto/muutl

StatFin. (2023b). Official statistics of finland: Population structure [ISSN = 1797-5359]. http://www.stat.fi/til/vaerak/index_en.html

StatFin. (2024). Travel-to-work areas. https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin___khi/statfin_khi_pxt_11xy.px/table/tableViewLayout1/

Sutela, H., & Larja, L. (2015). Ulkomaalaistaustaisessa väestössä paljon korkeasti ja paljon matalasti koulutettuja.

WID. (2018). World inequality database. https://wid.world/data/

A Serial correlation test

Table A1 presents Arellano-Bond tests for first (M1) and second (M2) order serial correlation in the outcome variables. The null hypothesis is no serial correlation.

Table A1: Arellano-Bond serial correlation test results

	Municipality		TTWA	
	M1	M2	M1	M2
Mean income	-9.59	-2.70	-1.41	-1.41
	[0.000]	[.007]	[0.159]	[0.158]
5^{th} percentile	-20.14	2.07	-4.82	-0.12
	[0.000]	[0.039]	[0.000]	[0.908]
10^{th} percentile	-16.24	0.024	0.298	1.84
	[0.000]	[0.981]	[0.766]	[0.065]
25^{th} percentile	-13.88	0.49	-0.14	0.85
	[0.000]	[0.625]	[0.891]	[0.395]
50^{th} percentile	-9.57	-3.59	-3.06	0.56
	[0.000]	[0.000]	[0.002]	[0.575]
75^{th} percentile	-13.30	1.25	-4.11	1.66
	[0.000]	[0.211]	[0.000]	[0.098]
90^{th} percentile	-14.56	-1.42	-5.50	-1.04
	[0.000]	[0.156]	[0.000]	[0.299]
95^{th} percentile	-15.11	-2.60	-3.49	-3.851
	[0.000]	[0.009]	[0.000]	[0.000]

Notes: The table test statistics from Arellano-Bond tests of first (M1) and second (M2) order serial correlation in the residuals of the IV estimates. The null hypothesis is no correlation and the test statistic is asymptotically normal. P-values in square brackets.

B First stage results

Table B1 presents full first stage estimates for the IV and multi-IV models of the results section. The first stage specification is presented in Equation B1 where ΔR_{lt-i} is the i^{th} lag ($i = 0, 10$) of the change in the immigrant-to-native ratio in location l at time t, Z is the instrument, γ captures time fixed effects and $\mathbf{X}$ is the familiar vector of controls. The ratio and its tenth lag are regressed against the instrument and its tenth lag, except for the first specification where no lags are included on either side.

$$\Delta R_{lt-i} = \gamma_{t-i} + z_{lt} + z_{lt-10} + \Delta \mathbf{X}_{lt-i} + \Delta u_{lt-i} \tag{B1}$$

Table B1: First stage estimates

	Municipal			Travel-to-work area		
	R	R	L10.R	R	R	L10.R
Instrument	0.966	0.869	-0.0155	1.144	0.595	-0.151
	(0.163)	(0.253)	(0.141)	(0.223)	(0.246)	(0.141)
L10.Instrument		0.0405	1.025		0.633	1.538
		(0.308)	(0.314)		(0.440)	(0.307)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	3597	1696	1477	612	288	252
F of instruments	35	14	18	26	16	15
Kleibergen-Paap χ^2	22.85	6.17		4.79	1.99	
	[0.000]	[0.013]		[0.029]	[0.158]	

Notes: The table presents the first stage estimation results. 'Ratio' refers to the ratio of immigrants to natives in the labor force - the explanatory variable of focus. The model is estimated in first differences, and controls and time fixed effects are included in each specification. L10 indicates the 10^{th} lag of the variable. Location clustered standard errors in parentheses.