Three Essays in Migration Economics

Ross McKenzie

A thesis presented for the degree of
Doctor of Philosophy

Department of Economics
University of Strathclyde
May 2022

Declaration

Acknowledgements

First of all, I am grateful to my supervisors and co-authors Marco Alfano and Graeme Roy for their continuous support during the PhD process. Although the path to get here was an unusual one they never failed to motivate and help me develop both professionally and as a person. This thesis would not have been possible without their endless patience and support.

I am also grateful to my co-author and friend Daniel Borbely. No matter the issue I knew I could turn to you for advice. Your constant enthusiasm and intelligent insights has made working together an extremely enjoyable experience and I look forward to that continuing into the future.

I would like to thank Orsolya Zombori and her colleagues at the Hungarian Central Statistical Office for their help parsing historical Census documents and to the UK Data Service for providing access to the secure Annual Population Survey, despite the difficulties in doing so during a pandemic.

I am also deeply indebted to those who helped support me while I waited to gain access to the secure data. Firstly to Markus for hiring me as a research assistant for two months to work on a variety of projects, although we only worked together for a short while I learned a lot in that time. To Robert Mitchell, Colin Carroll, Sophie Ellison and the rest of the Population and Migration team at the Scottish Government for giving me the opportunity to intern with them over this period and being so supportive and welcoming. Lastly, I would like to thank the University of Strathclyde and the Royal Economic Society for granting me the funding that made it possible to survive the pandemic and gave me the necessary time to receive the secure data that allowed me to produce this thesis.

I would also like to thank my PhD colleagues and friends Ali, Ayse, Ben, Chirsty, Gennaro, Geoff, Jon, Orion, Oyku, Piero, Sam, Sharada and Yuri for all the laughs and great conversations. I am also grateful to Jonathan, Stuart, Julia and everyone else from the Strathclyde Economics Department for their help, support and encouragement.

A special thanks to my parents Richard and Angie, my brother Cameron, my grandparents Richard, Marion and Magdalene, and all of my aunties. They are my greatest supporters and I could not have done this without their continuous love and support. Finally to my partner Yue, whether it was a joke to relieve my stress or discussing and picking apart new ideas, you have been there to support me everyday.

Synopsis

This thesis examines the impact of migration on economic development and native labour market outcomes. In three distinct essays, we make use of applied econometric methods to examine the consequences of forced out-migration on the development of townships in Hungary, the cross-occupational impact of immigration on UK native wages, and an in-depth review of common measurements of migration found in the literature. Our findings are intended to contribute to the wider migration policy debate using historical and contemporary evidence.

Chapter 1 is titled *"Forced Migration and Local Economic Development: Evidence from Post-war Hungary"*. This chapter investigates the persistent effects of forced migration on sending economies using the post-WW2 expulsion of German minorities from Hungary as a natural experiment. A large literature looks at the economic consequences of forced migration on receiving populations and the forced migrants themselves. Nonetheless, another interesting question, and one that is analysed less frequently is how forced migrations influence economic and social changes in the areas minority populations were expelled from.

The German minority population in Hungary have been present in the country for centuries and have developed their own distinct social, cultural, and economic structures in the localities they resided in whilst being largely assimilated into Hungarian society. The postwar forced migration was justified by political and economic motives which resulted in German populations, traditionally skilled in the agricultural and manufacturing activities required to cater to small economies, being replaced by lower skilled workers who were not well-suited to local labour markets.

In this chapter, we study the short and long-term effects of the forced migration of ethnic Germans from Hungary using both historical and contemporary data on roughly 2,000 Hungarian townships. Using historical census data from before and after the war we show that the township-level effectiveness of forced migrations is mostly explained by the pre-war shares of German minorities in each town and is unrelated to other factors that could induce selection bias in our analysis such as distance from borders or local economic factors. We then examine the effects of forced migration intensity on township-level aggregate economic outcomes, both short and long-term, through a continuous difference-in-differences framework, where the main explanatory variable is the forced migration intensity measure interacted with a post-migration indicator. To examine how the expulsions affected the local social fabric, we also examine the effects of forced migration on present-day trust levels. In most specifications, our empirical strategy controls for past (pre-migration) levels of outcomes and covariates in each township to ensure that results are not driven by the long-run spatial auto-correlation of economic variables.

We show that, while towns heavily affected by the expulsions were quite similar to other areas in terms of economic activity and labour market composition before the war, the forced migrations led to lasting reductions in economic activity, and an increasing reliance on agricultural labour. We further show long-term negative associations between forced migration and local trust levels, suggesting that the expulsion of Germans also affected the local social fabric. Our analysis reveals that forced migration can cause lasting regional inequalities in sending economies.

We present three main findings. First, the forced migrations can be associated with lower levels of economic activity (measured through population and labour force density) in heavily affected townships when compared to less affected ones. While these effects do diminish over time, they are still large and significant several decades after the expulsion of German residents. Secondly, we show evidence that the forced migrations can be associated with lower levels of trust in one's neighbours for individuals residing in affected townships. Finally, forced migrations led to permanent changes in local labour market composition: while the country's economy moved towards increasing the labour share of manufacturing workers, high forced migration townships increased their agricultural share, and this change persists into the current century. Overall, our findings show that forced migration can lead to lasting regional inequalities in sending economies.

This paper contributes to the small literature examining the effects of forced migration on sending (origin) economies by using administrative data from deportation registries to measure the intensity of forced migration at a highly granular township level and combining it with other historical sources to obtain more precise estimates of the extent to which each township was affected by the expulsion of its German residents. We further contribute to the literature on the lasting effects of historical events on trust. Finally, we contribute to the literature examining the effects of postwar German expulsions by being the first study to look at the effects of these expulsions in the context of Hungary.

Chapter 2 is titled *"Migration trends in Scotland compared to the UK and Why Measurement Matters"*. This chapter investigates the importance of how migration is measured when estimating the impact of immigration on native wages. The impact immigration can have on native wages is an important topic at the forefront of the migration policy debate in many countries worldwide. However, many studies have conflicting results. Card and Peri (2016) has shown that how migration is measured is important when interpreting results from the migration literature. Where if immigration is defined as the share of the local labour force, that I call the Immigrant Share, then estimates will be negatively biased as a result of relative-local demand shocks increasing native inflows and wages. I extend Card and Peri (2016)'s analysis by discussing three measurements of immigration: the Migrant Inflow, Immigrant Share, and the previously unexplored Migrant Native Ratio. Similar to Card and Peri (2016) I take first-order approximations of the Immigrant Share and in addition the Migrant-Native Ratio, a measurement commonly used in the UK literature. I show that in contrast to the Migrant Inflow, which is unbiased, both measurements of migration could cause a negative bias when estimating the wage impact of migration if native inflows and native wages are both positively correlated, which would be the case in the presence of relative positive local demand shocks. I then review key literature on the impact of migration on wages, where I show that different measures of migration are favoured across studies depending on the empirical specifications used and year it was published. To finish, I estimate the impact of immigration on native wages using the three different measurements of migration by utilising spatial variation across UK regions, where I separately regress native wages on the three different measure of migration. I account for endogenous migrant location choices by using the standard shift-share instrument first used by Card (2001). I find that, as expected, the Immigrant Share variable estimates a point estimate that is just under twice as negative as the point estimate when using the unbiased Migrant Inflow measurement. However, unexpectedly, the Migrant-Native Ratio produces a point estimate similar to the Migrant Inflow. I further show that although these results are not affected by native outflows as a response to higher migration, when regressing native outflows on the three measures of migration separately using Ordinary Least Squares, the Immigrant Share and the Migrant-Native Ratio estimate a more negative, although still insignificant, point estimate than the Migrant Inflow.

This chapter contributes to the literature in three ways. Firstly, it extends the analysis from Card and Peri (2016), to also consider the potential bias from measuring migration using the Migrant-Native Ratio and I am the first to analyse whether these different measures of migration produce different results when using UK data. Secondly, I provide an additional dimension to recent reviews of the literature on the impact of migration on native wages undertaken by Dustmann et al. (2016) and Edo (2019). Where I explore how the measure of migration has differed across three empirical approaches and the migration measurement choices have developed since Card and Peri (2016). Lastly, I contribute to the literature using a spatial approach to estimate the total effect of migration on native wages in the UK.

Chapter 3 is titled *"Cross-Occupational Effects of Immigration on Native Wages in the UK"*. This chapter estimates the effect of immigration into an occupation on wages of natives working in higher paid occupations, that we call cross-occupational effects. The impact of immigration on wages of natives remains a topic of intense debate, both in economic policy circles and the wider public discourse. The magnitude and even direction of the effect appears to vary based on the setting and approach taken (Dustmann et al., 2016). Studies typically investigate whether natives and migrants either compete with or complement each other in similar jobs or skills groups. There is a rich body of evidence looking at whether or not migrants either compete with or complement natives in the same part of the wage distribution. However, whether or not these same migrants yield benefits or costs to native workers just above or below them in the wage distribution has remained relatively unexplored. This paper contributes to the large, reduced form literature on the effect of immigration on native labour outcomes by allowing immigrants into one section of the labour force to affect natives in different occupations, this study provides a novel impact of migrants on native wages.

In this paper, we estimate the effect of immigration into an occupation on wages of natives working in higher paid occupations. To estimate the effect of immigration into an occupation on wages of natives working in higher paid occupations we use Office for National Statistics (ONS) data and divide workers in each of the 13 U.K. regions into 9 occupational categories based on the Standard Occupational Classification (SOC). To identify adjacent occupations, we first rank all 9 occupations according to the ordering provided by the SOC 2010 ordering with Managers, Directors and Senior Officials at the top and Elementary Occupations at the bottom. For each occupation o, we define the occupation below (o - 1) as the occupation with mean hourly earnings are one rank lower than o. Similarly, the occupation above (o + 1) is the occupation with mean hourly earnings one rank higher than o. Using these definitions, we regress yearly changes in native wages in occupation o on yearly changes the migrant-native ratio in occupations o, o-1, o+1. Following standard practice in the literature, we instrument migration flows using the supply-push instrument first detailed in Card (2001). Whilst we do not detect any meaningful effect of immigration within the same occupation-region group, we find that immigration into one occupation increases wages of natives working in the occupation ranked above by around 0.332 percent. Our findings are consistent with migrants increasing productivity and allowing natives to specialise.

This paper further contributes to the literature on mechanisms through which migrants affect native outcomes. To do so we provide suggestive evidence on potential mechanisms that can explain these cross-occupational effects. Firstly, we consider whether peer effects may impact productivity and therefore native wages because of social pressure to work harder and/or through knowledge spillovers. Using a simple regression analysis we show that that higher migrant inflows into an occupation are associated with higher levels of education in that same occupation, however we only find a weak positive association between increases in the average educational attainment of employees working in the occupation below, $o - 1$ and native wages in occupation, o. Secondly, following Peri and Sparber (2009) we consider whether migrants increase the wage of those natives in occupations above their own by allowing natives to specialise in better paid task in which they have a comparative advantage. We investigate this channel of impact by focusing on in-job training received by natives, where training is likely a pre-requisite for specialising in more complex tasks. However, although we find that migration into the same occupation , o, positively associates with the completion of in-job training of natives, migration into adjacent occupations provide no significant association. Our findings suggest that debates over the economic impacts of migration often ignore the potential spill-over benefits that a migrant can bring to the outcomes for native workers elsewhere in the wage distribution, particularly in lower wage occupations.

Contents

List of Figures

List of Tables

Chapter 1

Forced Migration and Local Economic Development: Evidence from Postwar Hungary [1]

[1]This chapter is co-authored with Dr Daniel Borbely, and began when we were both PhD students at the University of Strathclyde. It is based upon and extends research that appears in an earlier discussion paper found in Borbely and Mckenzie (2021)

1.1 Introduction

Forced migrations commonly target ethnic or religious minorities, whose replacement by other populations is often justified through political and economic motives. Recent examples include episodes of forced migration in countries such as Myanmar or Syria[2], while throughout history millions of people were uprooted as a consequence of ethnic cleansing, nation-building, and the partitioning of countries. A large literature looks at the economic consequences of forced migrations on receiving populations and the forced migrants themselves (see Becker and Ferrara, 2019). Nonetheless, another interesting question, and one that is analysed less frequently, is how forced migrations influence economic and social changes in the areas minority populations were expelled from. In this paper, we examine these effects using a historical natural experiment: the expulsion of hundreds of thousands of ethnic Germans from Hungary after the Second World War.

The German minority population in Hungary (also referred to as the 'Swabians') have been present in the country for centuries, and had developed their own distinct social, cultural, and economic structures in the localities they resided in, whilst being largely assimilated into Hungarian society (Spira, 1985; Toth, 1993; Fischer, 1992; Apor, 2004; Marchut, 2014; Seewann, 2012). The post-WW2 government justified their expulsion by political and economic motives (Toth, 1993; Apor, 2004; Seewann, 2012). First, the expulsions were propagated as a way to hold German minorities responsible for Hungary's wartime alliance with Nazi Germany. Second, due to a large number of Hungarian refugees in surrounding countries, population exchanges were encouraged by Allied governments. Finally, within Hungary, the communist postwar government saw the opportunity to reward their supporter base with the land and property confiscated

[2]Recent studies examining the economic effects of these events include Tumen (2016), Gehrsitz and Ungerer (2017), Balkan and Tumen (2016), Ceritoglu et al. (2017), Altındağ et al. (2020), and Segnana et al. (2020).

from German minority households, and thus raise the economic and social status of specific 'native' groups. Effectively, this resulted in the German populations, traditionally skilled in the agricultural and manufacturing activities required to cater to small local economies, being replaced by lower skilled workers who were not well-suited to local labour markets (Toth, 1993; Marchut, 2014).

To analyse the short and long-term effects of the forced migration of ethnic Germans from Hungary, we make use of both historical and contemporary data on roughly 2,000 Hungarian townships.[3] To assess the extent to which each township was affected by the forced migrations, we utilise Census data from before and after the war, along with administrative data on the number of Germans deported from each township. Our analysis shows that township-level variation in the 'effectiveness' of forced migrations is mostly explained by the prewar shares of German minorities in each town – the deportations were based on prewar Census information on people's ancestry – and is unrelated to other factors that could induce selection bias in our analysis (distance from borders or local economic factors, for example). We then examine the effects of forced migration intensity on township-level aggregate economic outcomes, both short and long-term, through a mix of panel and cross-sectional regression models where the main explanatory variable is a continuous measure of forced migration intensity. To examine how the expulsions affected the local social fabric, we also examine the effects of forced migration on present-day trust levels. In most specifications, our empirical strategy controls for past (pre-migration) levels of outcomes and covariates in each township, along with granular region fixed effects (at the township, district area, or county level) to ensure that results are not driven by the long-run spatial auto-correlation of economic variables (see Voth, 2021).

We present three main findings. First, the forced migrations can be associated with

[3]The sample of townships we have data on for our analysis fluctuates to some extent throughout the paper due to the fact that not all outcome data are available for all townships.

lower levels of economic activity (measured through population and labour force density) in heavily affected townships when compared to less affected ones. While these effects do diminish over time, they are still large and significant several decades after the expulsion of German residents. Second, forced migrations are associated with permanent changes in local labour market composition: while the country's economy moved towards increasing the labour share of manufacturing workers, high forced migration townships increased their agricultural share, and this change persists into the current century. We also provide some evidence, that, while only suggestive, indicates that this shift towards agricultural production cannot be associated with comparatively higher levels of agricultural productivity. We further provide suggestive evidence that these effects are potentially driven by those areas where native settlers (and not refugees) replaced German households. Finally, we show evidence that the forced migrations can be associated with lower levels of trust in one's neighbours and legal institutions (both measured in the present day) for individuals residing in affected townships, suggesting lasting changes in the local social fabric. Our paper also examines a number of other potential mechanisms driving long-term economic effects in addition to changes in the local labour force shares, such as changes in human capital at the local level, or changes in religious composition, but finds no evidence of lasting changes in these outcomes in response to the forced migrations. Our results, taken together with the evidence from anecdotal and historical sources, strongly suggest that changes in the composition of the local workforce is a likely factor behind the negative economic impacts of forced migrations on origin economies. Overall, our findings show that forced migration can lead to lasting regional inequalities in sending economies.

Our findings contribute to the small literature examining the effects of forced migration on sending (origin) economies. For example, Acemoglu et al. (2011) show that the expulsion of Jews from the Soviet Union led to smaller populations and lower wages in affected areas, while Arbatli and Gokmen (2018) show that historical Greek and Armenian minority presence in Turkey is positively associated with population density and human capital accumulation in the areas these minorities were expelled from. Other studies look at the effects of forced migrations on a diverse range of historical and contemporary outcomes, such as entrepreneurship (Grosfeld et al., 2013), banking (Pascali, 2016), literacy (Bharadwaj et al., 2015), and education (Akbulut-Yuksel and Yuksel, 2015). Our findings contribute to this literature by providing strong evidence that forced migrations have lasting negative impacts on origin economies, and identify changes in local labour force composition as a potential driver of these impacts. These findings are most similar to those in Chaney and Hornbeck (2016), who find that Spanish districts affected by Morisco expulsions failed to converge to other districts for nearly two centuries, and those in Testa (2021), who exploits spatial discontinuities in exposure to the expulsions of Germans from the Czech border areas to find that forced migrations led to reduced population density, higher unemployment, lower educational attainment, and smaller skills-intensive sectors. Our analysis yields similar results, but is unique in the literature in that it uses administrative data from deportation registries to measure the intensity of forced migrations at a highly granular regional (township) level. Using these data, in combination with other historical data sources, allows us to get more precise estimates of the extent to which each township was affected by the expulsion of its German residents. We also contribute to the literature on the lasting effects of historical events on trust (Nunn and Wantchekon, 2011; Grosjean, 2011; Becker et al., 2016), by providing evidence of long-term negative associations between forced migration and local trust levels. Finally, our paper contributes to the literature examining the effects of postwar German expulsions more specifically (Schumann, 2014; Braun and Kvasnicka, 2014; Semrad, 2015; Becker et al., 2020) by being the first study (to our knowledge) that looks at the effects of these expulsions in the context of Hungary.

The rest of this paper is organised as follows. Section 1.2 summarises the relevant historical background. Section 1.3 describes our data. Section 1.4 outlines our empirical strategy and summarises our results. Section 1.5 concludes.

1.2 Historical Background

1.2.1 German Minorities in Hungary

Germans have been present in the territories of historical Hungary for centuries. Their numbers vastly increased after the end of the Ottoman occupation of the Kingdom of Hungary near the end of the 17th century. Ottoman occupation and the wars to reclaim Christian territories left the country sparsely populated and with its labour force depleted. Organised immigration to replenish the labour force started in the 1710s and was subsequently promoted by three Habsburg emperors (Charles VI, Maria Theresia and Joseph II).[4] Migrant inflows were meant to increase the population of previously Ottoman controlled regions in Buda, Southwest Hungary, the Banat, and Szatmar county in Eastern Hungary (Apor, 2004). Most of the migrants were from the surrounding Habsburg and German territories, increasing the share of ethnic Germans within the Hungarian population. During the 18th century, Germans moved in through three large migration waves (or *Schwabenzug*, in German historiography), and after 1760 migrant inflows were not just encouraged, but also organised and financed by the Habsburg state apparatus, with help from existing migrant and family networks and specialised services such as recruiters, contemporary travel agents, and shipbuilders (Seewann, 2012). German migrant communities, later collectively referred to as Danube Swabians (Donau Schwaben[5]), remained an important part of Hungarian society, and maintained their cultural and linguistic traditions for the subsequent centuries (Toth, 1993; Apor, 2004).

[4] After the end of Ottoman occupation, Hungary fell under Habsburg rule. An account of this period is provided in Evans (2006).

[5] This term can be misleading given that a large number of German migrants came from territories other than Schwaben (Swabia, in present day Bavaria and Baden-Wurttemberg), see Apor (2004).

German migrants mostly (though not exclusively) moved into territories formerly occupied by the Ottoman Empire, where nearly 150 years of occupation left behind distinct socio-economic and demographic structures.[6] Indeed, our own data confirms that areas formerly under Ottoman occupation were more likely to have had a higher share of Germans prior to the forced migrations (see Table A.1). Ottoman territories were – already before German in-migration – much more ethnically diverse than the rest of the country, with a high proportion of residents of Romanian and Southern Slavic origins (Seewann, 2012). Incessant warfare and skirmishes by Ottoman and Habsburg armies in these territories also resulted in local populations becoming particularly mobile, with residents avoiding flat, arrable land, and moving into woodland areas for security reasons. When the German migrants moved into these territories, they brought with them modern agricultural tools to help revitalise local economies, but also introduced technological, social, and legal innovations to local communities (Seewann, 2012). The German migrants set up agricultural production on previously unused land, and built new settlements near agricultural areas.

During the 19th century, German inheritance customs, which prescribed that only the eldest son could inherit family land, led to younger children of German families either setting up land or businesses in nearby villages, or moving to larger towns or cities to enroll in intellectual occupations. This process has led to increased assimilation of German communities both spatially (see Figure 1.1) and in terms of their occupations. At the same time, tightly-knit township communities formed in rural areas where German and Hungarian families shared social and economic responsibilities, established shared

[6] These territories were invaded and annexed by the Ottoman sultan Suleiman the Magnificent during the 1520s to 40s. The Ottoman Empire occupied most of the central and southern parts of Hungary, consisting of more than 40% of the territory of historical Hungary. The north-western part of Hungary remained a part of Royal Hungary which recognised members of the Habsburg family as the Kings of Hungary. Eastern parts of the country became the Eastern Hungarian Kingdom and then the Principality of Transylvania, an Ottoman vassal state. For a summary, see Inalcik (2013). See also for a map of Hungary under Ottoman occupation see Magocsi (2002). For a map of the ethnic groups of Hungary before WW1 see Shepherd (1911) at Web Link.

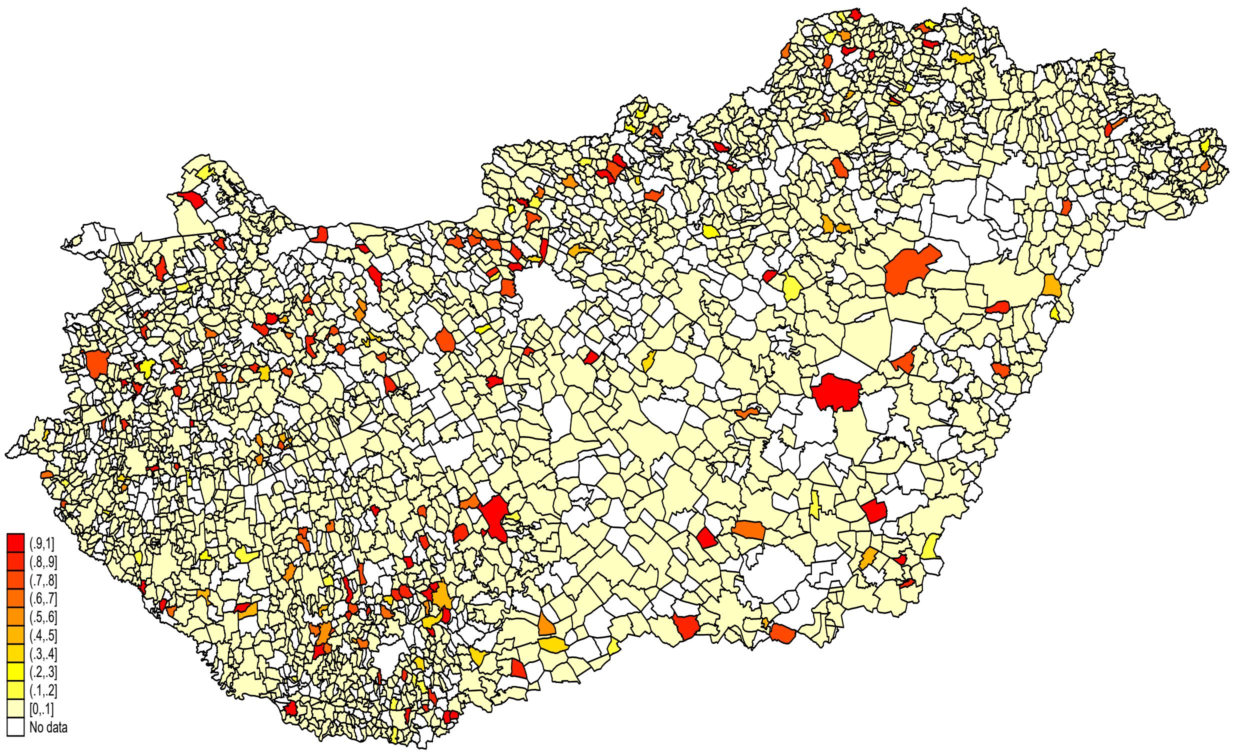
cultural traditions, whilst being fully tolerant of cultural (and linguistic) differences (Seewann, 2012). During the late 1800s, this idyllic state of multi-ethnic community existence was interrupted by the forced assimilation efforts of the increasingly nationalistic Hungarian authorities, who required minorities to adopt the Hungarian language and use it as the main language of instruction in schools. Many German families also had to change their last names to 'Hungarian' ones. Through forced assimilation, during the early 20th century, many German households effectively joined the middle classes of the emerging Hungarian nation state.

According to historical sources, prior to the expulsions in 1946-47, the German minorities in Hungary were very similar in occupational, political, and religious preferences to the 'Magyar' majority (Spira, 1985; Fischer, 1992; Marchut, 2014).[7] Younger members of German communities, who were enrolled in the Hungarian education system, were especially likely to have been assimilated (Marchut, 2014). In addition to the historical evidence, we are able to test the assimilation of ethnic Germans into the Hungarian economy. We assume, that if ethnic German have assimilated into the wider economy, then we would expect them to have similar economic outcomes to natives. Due to data limitations, we are unable to directly compare ethnic Hungarian and ethnic Germans in 1941. However, we can test whether the share of Germans present in a township effects the economic structure of these townships, where we use occupational shares to measure this. Our own analysis presented in Figure 1.2 largely confirms the assimilation of Germans to the majority population, at least as far as economic structures are concerned. Here, we regress the share of Germans in each Hungarian township area in 1941 on the local shares of different occupations. The figure shows that the share of Germans in townships was positively correlated with a higher share of helpers (domestic work-

[7] Although Germans were somewhat more likely to be employed in certain types of agricultural and manufacturing jobs, and were less likely to be employed in intellectual positions, most of these differences are due to the mostly rural locations of German majority townships (see Marchut, 2014), and are unlikely to lead to differences in the rural sample we investigate later on in this paper.

ers), and a lower share of physical workers, but there is no evidence of a correlation with occupational shares in key sectors.[8] Where, there is no significant difference in the Share of Agricultural workers and no significant difference in the Share of Manufacturing workers.

Figure 1.1. Share of Germans in Hungarian Townships in 1941 (Current Boundaries)



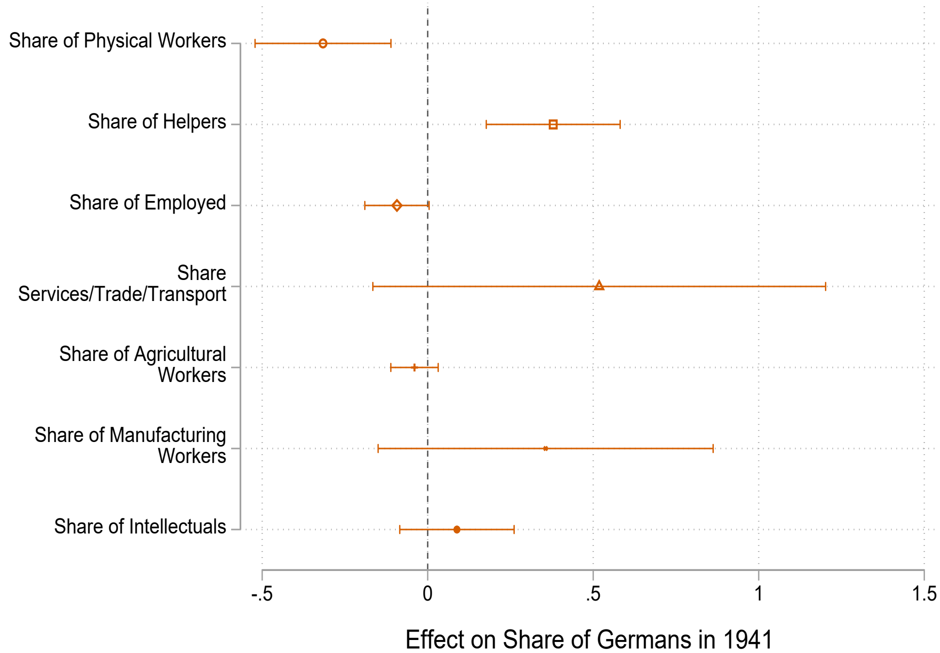
Notes: The map shows Hungarian township areas based on their current boundaries. Townships are matched to 1941 data on the share of Germans based on their names.

1.2.2 Population Transfers After World War II

By the end of WW2, there were almost exactly 500,000 people – roughly 5.6% of the country's population – of a 'German vernacular' living in Hungary, with 303,000 citizens

[8]We combined the services, trade, and transport sectors into a single measure due to very small numbers of people employed in each of these sectors in most townships. For example, transport workers have a mean share of 0.0058 (0.58%) and a standard deviation of 0.011. Including these variables separately does not change our conclusions but inflates the estimates and their standard errors.

Figure 1.2. Population Shares in Different Occupation Types and the Share of Germans in 1941



Notes: The point estimates are obtained from regressing the share of Germans in 1940 on the population shares of different occupations. The data on the right hand side variables is from the 1940 Census. Horizontal bars not spanning zero indicate significance at the 10% level.

claiming to be of German 'nationality' in the previous (1941) Census (Apor, 2004; Toth, 2018). In 1945, the country has emerged from the war on the losing side and conceded the territories gained in 1941.[9] These territorial changes, along with the multi-ethnic composition of the country's population have created multiple issues pertaining to minority and ethnic rights. On one hand, ethnic Hungarians now formed minority populations in the surrounding Slavic countries. On the other, both international and domestic political forces put pressure on Soviet-occupied Hungary to consider its treatment of its German minority population, mostly formed of families that have been residing in

[9]Hungary lost nearly two-thirds of its former territories after WW1. Some of these territories, including the southern parts of Czechoslovakia and Northern Transylvania, were reassigned to the country in 1938-39, as part of the first and second Vienna Awards. An account of this period is provided in Cornelius (2011).

Hungary for generations.

As the Red Army was advancing through Hungary during 1944, the country's German minority was starting to be deported to Soviet labour camps.[10] On December 22, 1944, the Red Army leadership ordered the 'full mobilisation' of German Hungarians to perform forced labour. The implementation of this order fell upon Hungarian local authorities. These actions were justified by the 'collective responsibility' of German minorities – most Germans living in Hungary were conscripted into the German army – for the German war effort and the associated war crimes committed by the Nazi regime.[11] The initial deportations were disorganised and poorly administered, inflicting serious damage on both German and Hungarian communities. In January 1945, Hungary signed an armistice with the Soviet Union that included a promise to hold war criminals to account along with the internment of all German nationals.[12]

Over the coming months, most of the Hungarian political elite embraced the position that a large-scale deportation of German minorities was needed. The official justification of this position usually revolved around the 'collective punishment' of Germans of all origins. Another objective of Hungarian policy makers was to free up land to repatriate Hungarian refugees from the surrounding countries, mostly from Czechoslovakia and Yugoslavia. According to Toth (1993), Hungary took in roughly 90,000 refugees from surrounding countries, including around 12,000 families from Slovakian territories. Around half of these families moved into properties confiscated from German households.

Aside from providing housing to refugees, freeing up land was also necessitated by the Land Reforms (*'foldreform'*) of 1945, whereby land was reallocated to the Hungarian 'proletariat' (see below). Historical sources also note that overpopulation in Hungary's

[10] An account of this period is provided (in Hungarian) by Toth (1993). An English language summary can be found in Apor (2004).

[11] Hungary was allied with Germany for most of the war and was occupied by the Nazi regime in 1944 after the Hungarian Government sought to negotiate a separate peace with allied powers.

[12] The armistice granted sovereignty to the nation of Hungary, although in practice this was limited to a large extent by the presence and authority of the Allied Commission.

major agricultural areas (the Hungarian Great Plain, for example) created the need to move the local peasantry somewhere else (Seewann, 2012). In practice, this meant that many belonging to the German minority were dispossessed of their land and properties, which were then redistributed to Hungarian settler populations. The Land Reforms thus led to plans to resettle roughly 180,000 people within Hungary, these settlers mostly moving into properties 'left behind' by German families (Toth, 1993). By the end of 1945, the Hungarian Parliament has created and accepted the legislation required for the deportation of almost half a million German minority citizens.[13] Citizens were deported if:

- They claimed to belong to the German minority; or claimed that German was their native language in the 1941 Census;
- They changed their name to a German surname during the war years;
- They were a member of the Volksbund or a German military group[14]

Some individuals, usually due to being married to 'native' Hungarians, were able to avoid deportations, however the maximum permissible proportion of these cases were capped at 10% of all deportation cases in each local authority.[15]

It is possible that there was some heterogeneity across Hungarian counties in terms of how 'effective' deportations were.[16] Deportations were said to be particularly intense in the Western border counties of Hungary (Gyor-Moson, Sopron, and Vas) to prevent territorial requests from neighbouring Austria, and were carried out more quickly in towns compared to urban areas (Apor, 2004). Beginning in 1947, pressure from Western countries and the US military prevented (and eventually halted) the deportations of the

[13] Executive order 12 330/1945. ME.

[14] The Volksbund was a pro-Nazi organisation of some of the German speaking minority in Hungary.

[15] Meaning at least 90% of the local German minority population had to be deported.

[16] We find no evidence of heterogeneity in deportation 'effectiveness' at the township-level in our empirical analysis below (see Section 1.4).

German minority. Overall, according to historical sources, roughly 230,000 (46% of the pre-deportation population claiming German nationality or ancestry) people of German ancestry remained in the country, although the majority of these people have had most of their land and possessions taken away from them by the Hungarian authorities (Toth, 2016).

While the deportations were partly to ’make space’ for the refugee populations returning to the country after WW2, political and economic considerations also played a significant role. Powerful factions in the Hungarian parliament of the time, such as the National Peasant Party (‘Nemzeti Parasztpart’, or NPP) or the Independent Smallholders Party (‘Fuggetlen Kisgazdapart’, or FKGP) also saw an opportunity to reallocate land from German peasants to their ’native’ supporter base. The NPP party leader, Imre Kovacs, put it thusly in a 1945 speech:

> We will deport the Swabians. It is not possible to have our best lands occupied by the Volksbund members, and for them to sprawl in their five-bedroom homes. There will be demand for these lands. The landless cottars of the Tiszantul[17], or the luckless Csangos who were kicked out overnight[18], should take the Swabian lands.[19]

With the deportations already mostly concluded, the social democratic wing (’Magyarorszagi Szocialdemokrata Part’, or SZDP) in the parliament raised concerns about various atrocities committed during the population transfers, which seemed to be mostly focused on the complete material dispossession of the German minorities. Additionally, there were concerns that the settler populations meant to replace the local Germans in their work activities were not able to do so whilst maintaining productivity (Toth, 1993; Marchut, 2014). For example, anecdotal evidence presented in the Hungarian parliament suggested that often times the settlers who replaced German households were

[17] The Transtisza region of Hungary.
[18] The Csangos are a Hungarian ethnographic group currently living in Romania.
[19] The speech was translated from the original Hungarian by the authors. Source: Web link.

only interested in selling (formerly) German property and possessions, and were often unskilled and tended to avoid work. According to Toth (1993), settler populations had trouble integrating into township communities in many cases, and were untrained in the agricultural and manufacturing activities required to cater to local economies, hampering productivity. Upon observing the way in which the deportations were conducted, the evangelical bishop Lajos Ordass summarised these issues thusly:

> We cannot just replace a group of people, who, through centuries built their own culture, religious life, and economic structures, and became deeply rooted in this land, without financial and moral consequences for the peoples afflicted. The same way as the century-old tree cannot be replanted without us causing its destruction. Our German families were prominent in agriculture, livestock breeding, craft manufacturing, and in general, in all economic activities. We cannot offset the losses incurred from expelling them through the resettlement process, which seems to be a troubling phenomenon to begin with. Miners and factory workers will not be able to replace those who conduct their trade properly.[20]

1.3 Data

For our analysis of the economic effects of forced migrations on sending economies, we collect data on a large sample of Hungarian townships. The administrative unit of 'township' is based on the designation (*'telepules'*, in Hungarian) by the Hungarian Central Statistical Office (KSH), and contains all cities, towns, villages, and smaller settlements, along with the surrounding land area within their administrative boundary limits.[21] Data collection requires matching township (or local area) level information across historical Census records and deportation registries (see below), resulting in

[20] The quote was translated from the original Hungarian by the authors. Source: National Evangelical Archives (Evangelikus Orszagos Leveltar), Church Records (317/1946), available at: Web Link

[21] See the KSH website for more information on administrative units: Web Link. In the English translations of recent publications, these territorial units were referred to as 'settlements'.

samples of townships for whom outcome data are continuously available, which, for some outcomes, excludes some larger urban areas. [22] Below we detail our various data sources and the sample selection process.

Data on forced migration intensity. Information on the number of Germans in each township before the forced migrations is available through the 1941 Census. We consider the number in each township who claimed that their mother tongue was German.[23] We use this designation because this was the criteria – having German nationality or German as a native language – used to identify the Germans to be deported by Hungarian authorities during the expulsions. Our final sample of townships has in total 221,186 German inhabitants according to the 1941 Census, meaning that our sample contains roughly half of the country's estimated German minority population at the time.[24] To determine the number of Germans in each township after the deportations we rely on two main data sources. First, we use data on the number of Germans in each township from the 1949 Census. These data are however likely to underestimate the true number of Germans, who had a strong incentive to deny German nationality (or ancestry) in the postwar Census after witnessing the deportations a few years prior. To overcome this issue, we make use of the Registry data on the deportations compiled by Hungarian authorities.[25] These data indicate the number of Germans (tracked using the 1941 Census records) in each township who were deported, killed in the war, or were captured and ended up in POW camps. By subtracting these numbers from the 1941 Census data we can get a more realistic estimate of the number of Germans in each township after the

[22]Data on certain large urban areas were not reported consistently throughout various censuses due to changes in city and administrative boundaries. For example, the territory of Budapest now includes a number of townships and small cities that used to be separate administrative entities in prewar censuses, and is usually disaggregated into smaller districts whose boundaries also changed during our sample period. To address concerns related to sample selection, we carry out several checks below.

[23]During the 1941 Census, there was an incentive to report German ancestry as Hungary was allied with Germany in the war.

[24]This number is estimated based on historical sources, see Apor, 2004.

[25]The registry data are available here: Web Link.

deportations.[26] Overall, in our sample of townships, the 1949 Census data suggests that roughly 80% of the pre-war German population was no longer present after the deportations, while the Registry data shows that this share is around 58%, the latter number likely being a more accurate estimate of the true share of those no longer residing in Hungarian townships.

To gauge to what extent different areas were affected by different settler populations replacing German households, we also collect information on 1) the number of Hungarian settlers moving into (or planning to move into) each destination county through land reform policies in 1945-46 and 2) a sample of refugees from Slovakia moving into different Hungarian counties in 1947. These data are from the analysis of the resettlement records by Toth (1993). The Slovakian refugee sample consists of approximately 3,000 families, but contains only those families who moved between April and June, 1947, and therefore is likely not representative of the destinations of the entire Slovakian refugee population of around 12,000 families. These data are based on county-level aggregates from the Hungarian Ministry of Interior Affairs and are likely to be imprecise as they generally fail to track settlers after entering them into the resettlement registries (Toth, 1993).[27]

Data on economic outcomes. We use a number of different data sources, both historical and contemporary, to measure the effects of the forced migrations on various economic outcomes. First, we use the 1949 Census to collect township-level data on employment rates and the shares of local residents employed in different sectors.[28] To examine present-day outcomes, we collect the same data using the most recent Census,

[26]According to Marchut (2014), the Registry data does likely contain some errors, but is still mostly reliable and constitutes the best source of information on the numbers deported from each township.

[27]Destination counties in the settler destination data are based on old county ('varmegye') borders, which we map into present-day counties based on the (current) county each township in our data set is located in. The mapping is based on information from **?**.

[28]Note, that the 1941 and 1949 Census data are only available in PDF formats, which we parsed and exported into MS Excel using parsing algorithms. We checked the parsed documents afterwards for missing values. Data does not seem to be collected on many townships in the 1949 Census which leads to a reduced sample size of townships for which we have data available from both Census years.

conducted in 2011. Second, we use township-level Census data on population density from 1920 to 1990, collated in (roughly) ten-year intervals that correspond to each Census wave. These longitudinal data were published through the 1990 Census, and are therefore available for a larger sample of townships when compared to outcomes where we relied on matching townships from parsed historical Census documents. Following several studies in historical economics (Davis and Weinstein, 2002; Bleakley and Lin, 2012; Alsan, 2015), we use population density as a proxy for local levels of economic development. As an additional outcome measure, we also use the local employment rate (from both 1949 and 2011) to examine changes in economic activity levels. Moreover, we complement our analysis with a number of additional data sources on potential outcome variables. We collect township-level Census data on educational outcomes and religious shares from 1941 and 2011, along with data on agricultural outcomes from the Hungarian Central Statistical Office (KSH) from 2007. Finally, we collate agricultural productivity data for different crops from the FAO GAEZ database for Hungarian district areas ('jaras') for the year 2010.[29] Agricultural productivity is measured using current crop yields. These are distinguished into labour intensive and capital-intensive crops. This allows us to capture two potential outcomes. Firstly, is there an association between forced migration intensity and crop yields. Secondly, is there an association between forced migration intensity and the use of capital-intensive crops.

Data on trust. We make use of the Life in Transition Survey (LiTS) conducted by the European Bank for Reconstruction and Development (EBRD) to measure various forms of trust for residents of Hungarian townships.[30] We use the third and latest round of the LiTS survey, which was conducted in 2016. The LiTS contains rich information on socio-demographic measures along with various measures of trust in public services for

[29]The FAO GAEZ database is available online through the FAO website (Web Link).

[30]Previous studies that used these data to measure trust in an economic context include Grosjean (2011) and Becker et al. (2016). The survey can be accessed through the EBRD website (Web Link).

a representative sample of 1,500 Hungarian households, located in 62 township areas.[31] Respondents indicate the level of trust (on a scale of 1 to 5) they have in various entities, including the national government, courts, and neighbours. We use this information to create dummy variables corresponding to each of these categories, where a value of one indicates at least some level of trust. We match the individual-level trust data to our forced migration intensity data using information on the township of residence for each individual in the LiTS survey.[32]

Data on covariates. We also collate data on a number of key demographic variables and township-level aggregates which we use as controls in our regression models. We use the 1941 Census to collect data on the demographic (age, gender, education, etc.) composition of each township, along with local labour market characteristics (shares of different occupations, share of workers employed, and so on). We also collect information on various local characteristics, such as infrastructure (road and rail networks), and geographic variables (suitability for cultivation of crops, size of arrable and non-arrable land areas, distance to country borders). Data on agricultural outcomes are from the Hungarian Agrarian Census of 2007 conducted by the Hungarian Central Statistical Office, while distance data is calculated using (current) township centroid coordinates in QGIS. Some of these measures are time-invariant and therefore constant across sample years.

Sample selection. Our analysis below uses three main samples of townships: 1) the full sample of townships for which information on population density is available from 1920 to 1990 through the 1990 Census; 2) the sample of townships that we can match across the 1941 and 1949 Censuses and 3) the sample of townships that we can match across the 1941 and 2011 Censuses. In all of these cases we match with the Registry data on the

[31] The survey is collected in 29 countries, but naturally we only use the Hungarian sub-sample for our analysis.

[32] Note, that this includes a smaller (but representative) sample of townships when compared to our main analytical sample.

forced migrations which may not contain all townships in the specific Census sample.[33] To what extent do these data cover the pre-war German population of the country? The 1920-90 sample contains 92% of the 'Germans' from the 1941 Census (see above), the townships in the 1949 sample contain roughly 31%, while the townships in the 2011 sample contain around 73%.[34] To alleviate concerns that our three main samples of townships are selected ones in that they contain (or do not contain) townships based on their initial German shares, we regress the share of Germans in 1941 for the full 1941 Census sample on a dummy variable indicating those townships that are dropped from the sample once we match to data from years after. Our results for all three main samples are reported in the Appendix, Table A.2. These results clearly indicate that there is no significant correlation between initial German share and a township being dropped from the sample. In Appendix Table A.3, we also check how similar the selected and not selected townships are in each sample using the (few) variables we consistently have information on (1941 population, area, and urban/rural status). In all cases, the selected sample contains (on average) more rural, smaller and less populous townships, but these differences are not always statistically significant.[35]

[33]The Registry data also contain the number of Germans in each township according to the 1941 Census as this was the criteria used to deport Germans in each township.

[34]These numbers are based on the numbers of Germans each township in each sample *had* in 1941 based on Census data. Determining the actual total number of Germans is a difficult task. As noted in Section 1.2, the country's estimated German minority population at the time was around 500,000 individuals, while around 300,000 claimed to be of German 'nationality' in the 1941 Census. The total number estimated based on historical sources and is based on the number of people with a 'German vernacular', see Apor, 2004. It is worth noting though that, after centuries of co-existence, 'German' (Schwabian) identity was highly malleable and in many cases perfectly compatible with also having a 'Hungarian' identity. The actual number of 'Germans' - if such a number is actually possible to quantify - is therefore very difficult to determine and depends on the criteria used to establish 'German' identity.

[35]This is also due to the fact that Budapest, the country's largest and most populous city by far, is excluded from all three main samples (see above). For example, the 1920-90 sample is fairly unbalanced when compared to the excluded townships, but very few townships (41) are excluded from this sample and the presence of Budapest in this group skews the means. Once we remove Budapest the two samples are balanced so it can be said that this sample is representative of townships outside of Budapest. The Hungarian capital is an outlier in terms of its population and levels of economic activity and has experienced a sustained period of growth both in terms of its area and population in the postwar decades (Brown and Schafft, 2002). Moreover, numerous townships merged into Budapest over the sample period, which makes it difficult to collect consistent outcome data on the city (and its surroundings) over time. The phenomenon of a single large city that is an outlier is not uncommon in the world. For example in countries such as Japan, New Zealand, Peru, Argentina, Chile, Lebanon and Greece.

Summary statistics for our 1949 and 2011 Census samples are shown in the Appendix, Table A.4 and Table A.5, while Figure 1.1 maps the pre-migration share of Germans across the sample of townships for whom this information is available.[36]

1.4 Empirical Evidence

In this section, we examine the effects of the postwar forced migration of German minorities on past and current economic and social outcomes. As our main explanatory variable, we use the Registry data (see above) on forced migrations to calculate the population adjusted difference between the pre and postwar number of Germans in each township, constructed as:

$$FM_i = -\frac{(Germans_{1949} - Germans_{1941})}{Germans_{1941}} x \frac{Germans_{1941}}{Population_{1941}}$$

where FM_i is the cross-sectional forced migration intensity for each township 'i'. The minus term at the beginning of the formula is included so that larger positive numbers indicate higher levels of forced migration intensity. This measure considers forced migration intensity directly and is based on the formula outlined in Lee et al. (2017). As a robustness check, we consider two alternative measures of forced migration intensity in Section 1.4.3 below.

The key issue when it comes to identifying the economic effects of the expulsions pertains to whether forced migration intensity was exogenously determined or whether it was determined by local factors. If certain economic or geographic factors influenced the effectiveness of deportations as well as our outcomes, then our estimates would

[36] This map considers all present day townships for which we have information on spatial boundaries. Naturally, the 1941 Census, for which we have information on pre-migration German shares, does not contain all of these townships.

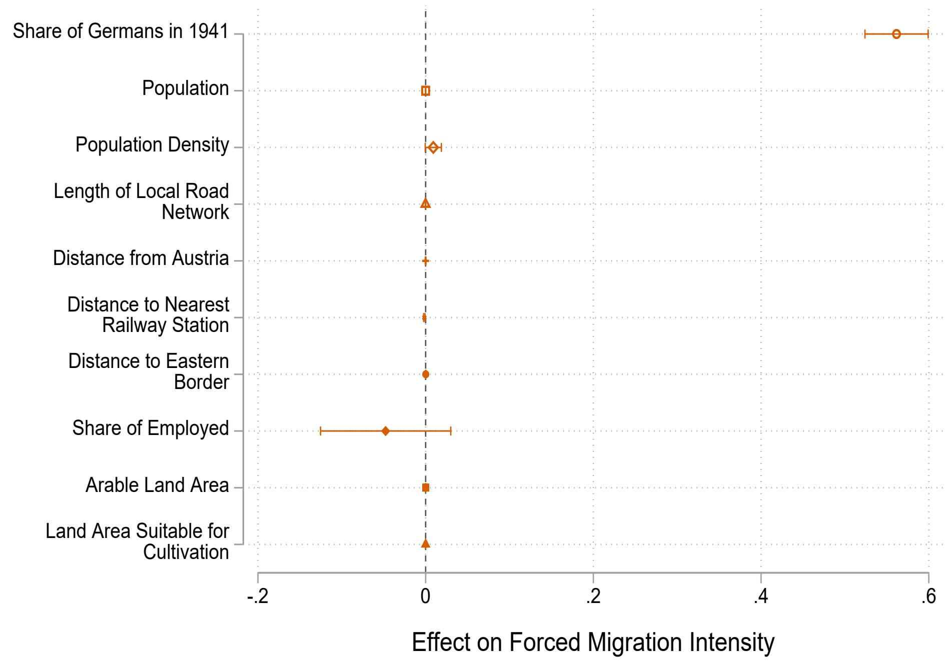
likely be biased. Indeed, as mentioned in Section 1.2, some historical sources suggest that forced migration intensity was higher in some areas, for example, near the Austrian border, or in smaller townships. It is also possible that, due to the effort to replace the expelled German households with refugees and politically loyal natives, towns that were more 'attractive' to settler populations may have experienced a more efficient deportation process.

To test for the presence of factors correlated with forced migration intensity, we regress the intensity measure (based on the Registry data) on a number of key variables that could have influenced the efficacy of the deportation process. First, to check whether deportations were more effective in larger towns or more densely populated areas we include prewar township population and population density. Second, to assess whether the local rail or road infrastructure made deportations easier, we include the length of the local road network along with each township's distance to the nearest railway station. Third, to examine regional patterns in forced migration intensity, we include the distance from the Austrian border, to check if deportations were indeed more efficient in areas closer to this border (see Section 1.2); and the distance from the Eastern borders, to see if deportations were more (or less) effective in territories that the Red Army initially occupied. Fourth, we include the prewar share of employed residents in each township to check if deportations were more effective in towns with a more attractive labour market. We also include the size of the local arable land area, and the size of the local land area suitable for agricultural cultivation, to check whether the potential for agricultural productivity in the local area influenced forced migration intensity. Finally, we include the share of Germans in 1941 as the most likely factor influencing the extent to which each township was affected by the forced migrations. The point estimates corresponding to each of these variables are plotted in Figure 1.3. For most of these variables, with the exception of the pre-migration share of Germans, we find precisely estimated null effects on forced migration intensity, suggesting that the majority of the factors listed did not influence the effectiveness or intensity of the expulsions to any extent. In Figure 1.4 we also plot the residuals from regressing forced migration intensity on the share of Germans in 1941. The residual plots show that there is little variation left in the forced migration variable on top of what the initial share of Germans already captures, and this is even the case for townships that had a non-zero initial share of Germans (Panel B).[37] These figures tell a clear story: the main (and only) factor significantly correlated with forced migration intensity is the initial share of Germans in each township, and there is limited variation in the extent to which towns with German residents prior to WW2 were affected. We show visual evidence of this correlation as well in the Appendix, in Figure A.1.

The findings presented in Figure 1.4 also suggest that high and low forced migration townships are very similar in terms of observable geographical and population characteristics. If we assume the initial share of Germans – mostly determined by the geographic patterns of Ottoman occupation and subsequent spatial assimilation of German minorities (see Section 1.2) – to be exogenous, then, conditional on pre-existing differences in economic outcomes, we can consider our measure of forced migration intensity to be appropriate to assess the causal effects of the expulsions. Recall from Figure 1.2, that the economic and labour market structures of 'German' towns were largely similar to other townships, and our sample of townships is balanced in terms of observable (pre-migration) economic characteristics. In settings where we can control for pre-existing differences in outcomes, we therefore attribute any post-migration divergence in outcomes across townships differently affected by forced migration to the expulsion of German minorities.

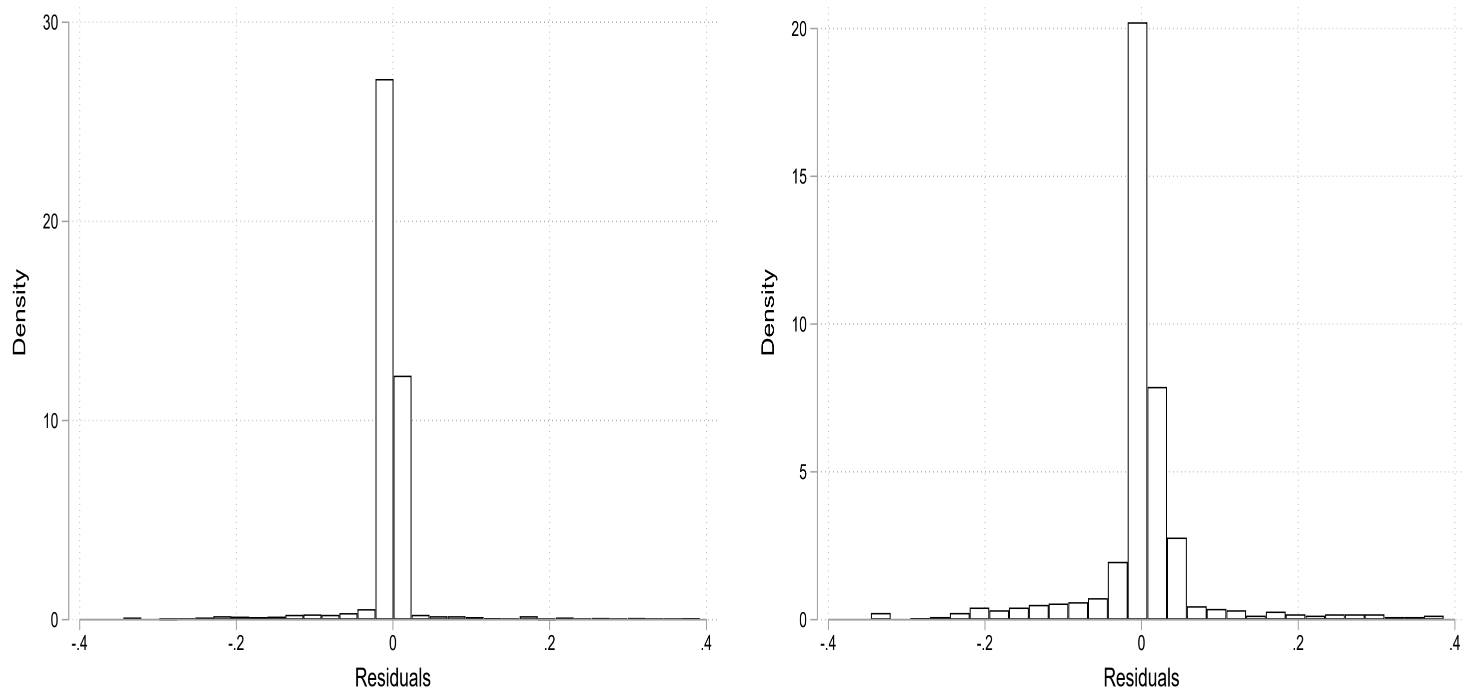
[37]Naturally, our measures of forced migration intensity are all to some extent dependent on the prewar shares of Germans in each township. Nonetheless, if, for example, several townships with high prewar shares of Germans would have experienced little (or no) forced migration, this would still lead to variation in forced migration intensity that is unrelated to the initial share of Germans. The fact that we do not observe such variation in the residuals suggests that no other factors likely influenced the effectiveness of forced migrations at the township-level.

Figure 1.3. Predictors of Forced Migration Intensity



Notes: The point estimates are obtained from regressing our main measure of forced migration intensity on the variables on the left hand side of the figure. Data on geographical distances are from GIS shape files. Data on arable land area and land area suitable for cultivation are from the FAO GAEZ database.

Figure 1.4. Residuals from Regressing Forced Migration Intensity on the Share of Germans in 1941



(a) Full Sample - All towns

(b) Towns with a non-zero share of Germans (1941)

Notes: The measure of forced migration intensity here is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. Residuals are obtained from regressing forced migration intensity on the share of Germans in 1941. Panel (a) shows the residuals for the full sample of townships. Panel (b) shows the residuals for a sample of townships where at least some level of forced migration took place.

To assess the effects of forced migration intensity on township-level economic outcomes, we rely on cross-sectional differences (across townships) in forced migration intensity as our identifying variation. The following sections examine the effects of forced migration on various outcomes.

1.4.1 Effects on Population Density

We begin by looking at the effects of the forced migration of German minorities on the population density of affected townships. For this outcome, we use a panel of the full sample of townships, measured at approximate ten-year intervals between 1920 and 1990.[38] Population density is measured by dividing a township's population by its total area (measured in square kilometres). We estimate the following two-way fixed effects (TWFE) regression model:

$$y_{it} = \alpha + \gamma \ x \ FM_i \ x \ Post_t + \theta_i + \theta_t + \theta_{ct} + \epsilon_{it} \ (1)$$

where our main coefficient of interest is γ, which measures the effect of forced migration intensity on our outcomes of interest (y_{it}). The term $Post_t$ indicates the year(s) after the deportations, taking a value of one in or after 1949. This specification follows a continuous difference-in-differences approach (see Card, 1992), where outcomes are compared across low and high treatment intensity townships, 'i', over time, 't'.[39] Township fixed effects 'θ_i' should control for all time-invariant factors that are fixed at the township level, while 'θ_t' are Census year fixed effects. We also add year fixed effects specific to each county ('θ_{ct}') to control for time-varying regional shocks. We mainly exploit

[38] To be precise, the '1940' data is from the 1941 Census while the '1950' data are from the 1949 Census.
[39] We summarise the limitations of this approach in Section 1.4.3 below.

cross-sectional differences in forced migration intensity across townships, and therefore a risk is that most of the effects we unveil are due to economic geography, whereby stable patterns in certain regions are correlated with (or causing) long-term outcomes whilst also being correlated with the explanatory variable itself.[40] To prevent spatial auto-correlation of outcomes from driving our results, we include town fixed effects in all specifications and report p-values for Moran's I statistic for our main specification below. Standard errors are clustered at the township-level in all subsequent specifications. Table 1.1 reports the results from an OLS regression using Equation (1), where the outcome is township-level (log) population density at the end of each ten-year interval from 1950 to 1990. Following Acemoglu et al. (2011) we also include ten-year lag of (log) population density for each of these years after 1950 to ensure that results are not predicated on pre-existing (but post-reform) trends in township-level population density. For example, if we did not include this term, then our point estimate for 1970 would be the pooled effect of 1950, 1960 and 1970. However, by including ten-year lags, we control for the trends in log population density post-treatment and before 1970, such that the point estimate is the marginal long-term impact of forced migration that occurred between 1960 – 1970.

Our results indicate that forced migration had a persistent negative effect on population density in origin townships, though this effect does fade over time. The largest effect is observed in 1960, a 24.3% (0.47 SD) reduction in population density associated with a unit increase in township-level forced migration intensity. This effect reduces to a 11.5% (0.19 SD) reduction by 1990. We report p-values for Moran's I statistic, which measures the extent to which spatial autocorrelation explains the values of our outcome variable. The reported p-values are between 0.10 and 0.25, suggesting that spatial autocorrelation in population density is unlikely to drive our results.

[40]For an overview of this issue in the literature on the persistent effects of historical events, see Voth (2021).

Having a longer time-series from 1920 to 1990 also allows us to check for the existence of pre-trends in population density across high and low forced migration intensity townships. Pre-migration divergence in time trends could shed doubt on our identifying assumption that differences in outcomes across townships can be contributed to the forced migrations. Panel a) of Figure 1.5 shows time trends in (log) population density for the top tenth percentile of townships in terms of forced migration intensity against towns with no forced migration (the median level of forced migration in our sample of townships is zero). It is evident from Figure 1.5 that the two groups of townships followed largely parallel trends in (log) population density up until the deportations, diverged during the two decades afterwards, and ended up on a negative trajectory after the 1960s. This overall downward trend in later decades is most likely due to increased urban migration within Hungary (Brown and Schafft, 2002).[41] This trend is robust to how we define high forced migration intensity townships, where Figure A.2 shows similar pre-trends for any forced migration, and the top 10th, 20th and 30th percentile. Where expectedly, the less forced migration intensity townships we include in our definition, the smaller the fall in population density in 1950.

We also assess the longer-term effects of forced migration on population density, along with the existence of pre-trends, through an event-study specification in panel b) of Figure 1.5. In this specification, we interact our forced migration intensity measure with year fixed effects. Point estimates correspond to differences in log population density between high and low forced migration intensity towns, relative to the same differences in a reference year (1940), for each year of the sample. The results in Figure 1.5 confirm the strong negative effect on population density and that these effects do persist over time. Point estimates from the pre-period are close to zero and are not significant in 1920, and only marginally significant in 1930, suggesting that our results are not predicated on pre-existing differences in trends across high and low forced migration townships. Where

[41] As outlined in Section 1.3, our sample mostly consists of rural townships.

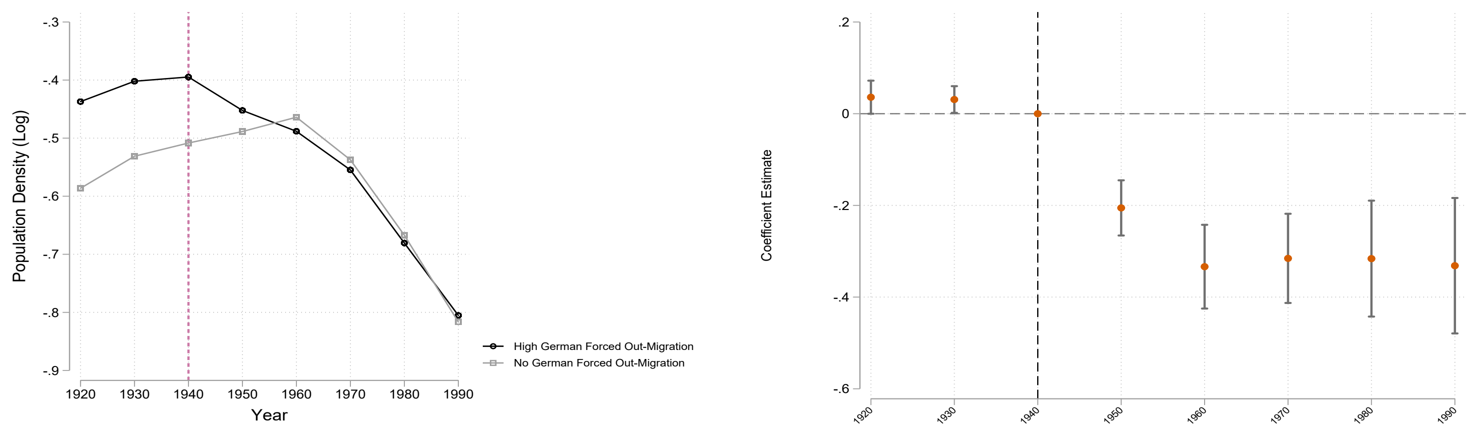
the coefficient itself is extremely small. The estimated coefficient for 1920 is 0.036 while for 1930 it is 0.031, but the minimum value with a 95% confidence interval is .0019. This shows that although the 1930 estimate is marginally significant at a 5% level, it is still smaller than the 1920 estimate. It also shows that the point estimates for 1930 are very small and close to zero, especially when compared to the subsequent post-treatment effect, dropping to around -0.2 in 1949 and even further in subsequent years. Finally, Appendix, Figure A.5 shows that when using two alternative measures of forced migration intensity, the pre-trends are clearly insignificantly different from zero. Overall, the results in this section provide strong evidence that the forced migrations led to lasting regional differences in population density, though our baseline results suggest that these differences did diminish over time. A possible explanation for the convergence we see starting around the 1970s is that the introduction of Hungary's New Economic Mechanism – a set of market-oriented economic policies that limited the role of central planning to some extent (see Balassa, 1983) – eliminated some of the regional inequalities by increasing the efficiency of labour markets. Another possibility is that the increasingly urban concentration of the Hungarian economy (Brown and Schafft, 2002) eventually led to a decrease in population for all rural township areas.

Table 1.1. OLS Results - Population Density Over Time

	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Forced Migration Intensity X Post	-0.221***	-0.243***	-0.183***	-0.156***	-0.115***
	(0.030)	(0.031)	(0.028)	(0.031)	(0.027)
Observations	8770	11614	14505	17380	20236
R^2	0.970	0.947	0.933	0.935	0.942
Moran P-value	0.10	0.12	0.25	0.14	0.19
Mean DV	-0.48	-0.47	-0.47	-0.49	-0.52
SD DV	0.51	0.52	0.54	0.57	0.60
Town FE	Yes	Yes	Yes	Yes	Yes
County x Year FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: The point estimates are obtained from regressing the log of population density on our preferred forced migration intensity measure, which is the population adjusted measure relying on Registry data. Each column shows the effect of the forced migrations in a different sample year. In each column specification, all preceding years are included in the sample, while all subsequent years are excluded. All specifications include township fixed effects, county times year fixed effects, year fixed effects, and the ten-year lag of the dependent variable. Observations are the number of township-years. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Figure 1.5. Event Studies - Population Density Over Time



(a) Population Density (Log) Over Time

(b) Treatment: Forced Migration Intensity (Registry Data)

Notes: In panel a), the measure of forced migration intensity is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. The black connected line with the circles plots (log) population density over time for high (top 10th percentile) forced migration townships. The grey connected line with the squares shows the same outcome for townships where no forced migration had taken place. For panel b), the point estimates plotted are from a version of Equation (1) where year fixed effects are interacted with the forced migration intensity variable. The reference year is 1940. Confidence intervals (vertical bars) not spanning zero indicate significance at the 5% level.

1.4.2 Mechanisms

Our results provide convincing evidence that the postwar deportations of German minorities had a lasting impact on local economic development. In the following section, we attempt to explore plausible mechanisms through which these changes materialised. For the following outcomes, we have data on outcomes from only a before and after period, and estimate two-period specifications. In all two-period specifications, our main explanatory variable (forced migration intensity) is cross-sectional, we therefore collapse our data into a cross-section of townships with outcomes measured in the post-migration period and covariates (including past levels of outcomes) measured in both pre and post-migration periods. Following Arbatli and Gokmen (2018), we estimate a regression model of the form:

$$y_i = \alpha + \gamma \ x \ FM_i + \beta^{`} X_{i,1940} + \delta^{`} X_i + R_i + \epsilon_i \ (2)$$

where y_i is a post-migration outcome of interest in township 'i', and FM_i measures cross-sectional differences in forced migration intensity across townships. We include pre-treatment economic controls (population density, shares in different occupations) through the term $X_{i,1940}$ to proxy for initial economic conditions at the township-level, along with exogenous controls X_i to account for possible sources of persistent differences in outcomes. The term R_i are region-specific fixed effects associated with the present-day county (or district area) of each township. These fixed effects are for all nineteen Hungarian counties in most specifications, or for all 174 district areas ('*jaras*') in some specifications using present-day outcomes. The average county contains approximately 146 townships in our sample, while the average '*jaras*' contains roughly 19. These region fixed effects should control for all time-invariant factors that are fixed at the regional level, for example the persistent effects of Ottoman occupation (see Section 1.2). Standard errors are clustered at the township-level in all specifications. In most specifications, we include pre-migration levels of our outcome variable to control for pre-existing level differences in outcomes across townships. In these specifications, conditioning on pre-migration differences in outcomes allows us to estimate two-period (pooled) difference-in-differences models. In specifications where we do not have information on pre-migration outcomes we uncover historical correlations (see Section 1.4.2.5 below). The following subsections consider various potential mechanisms and outcomes of interest.

1.4.2.1 Changes in the Local Employment Rate

While our results in the previous section suggest a clear negative effect on population density, we cannot be certain that these effects are due to changes in economic activity (for which population density is a proxy), and not due to changes in migratory and resettlement patterns. For example, it is possible that settlers did not replace Germans in the same numbers, or that they left their new residence soon after the expulsions, leading to lasting changes in population density as affected townships were never repopulated (Toth, 1993; Marchut, 2014). Naturally, these lasting negative effects on township populations would likely have a detrimental effect on local markets as well. Nonetheless, forced migration can affect local economies more directly by changing the composition of the local labour force and changing the availability of workers and skills at the township level. To assess economic effects further, we estimate Equation (2) using the township-level employment rate as the outcome variable. We conduct this exercise using both the employment rate in the short-run (in 1949), and on the long-run (2011), as outcomes. The results are summarised in Table 1.2 and Table 1.3. The results presented in these tables provide evidence of a clear negative short-run effect on local employment rates, this effect however does disappear by 2011, and how long the initial effect persisted remains unclear.[42] Note, that estimating the same two-period specification on the same sample of townships, but using (the log of) population density in 2011 as the outcome confirms the lasting negative effects of forced migrations on population density (see Appendix Table A.6). A possible explanation for the divergence in lasting effects on population density compared to the employment rate is that while the local population, and possibly the labour force (see below), shrunk due to the expulsions, the rate of employment eventually returned to a steady-state as local labour markets adjusted to the influx of the new labour force. We detail such changes in the composition of the labour force as a potential channel for long-term economic changes in the next section.

Table 1.2. OLS Results - Employment Rate in 1949

	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Registry Data)	-0.089	-0.157***	-0.098
	(0.055)	(0.054)	(0.065)
Number of Townships	799	664	664
R^2	0.005	0.099	0.130
Mean DV	0.45	0.44	0.44
SD DV	0.16	0.16	0.16
Covariates	No	Yes	Yes
County FE	No	No	Yes

Notes: The point estimates are obtained from regressing the employment rate in 1949 on our preferred forced migration intensity measure, which uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects and all our covariates. Covariates include the 1941 labour market shares of different sectors and pre-migration employment rates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

1.4.2.2 Changes in the Composition of the Local Labour Force

Historical sources suggest that differences in skills between German and settler populations led to fundamental changes in the composition of the local labour force after the expulsions of Germans in certain townships (Toth, 1993). In this section, we test this

[42]Unfortunately, we do not have the same panel data available for employment rates as we do for population density.

Table 1.3. OLS Results - Employment Rate in 2011

	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Registry Data)	0.004 (0.021)	0.021 (0.022)	0.013 (0.021)
Number of Townships	1270	1270	1255
R^2	0.385	0.412	0.545
Covariates	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Labour Shares in 1941	No	Yes	Yes
Area FE	No	No	Yes

Notes: The point estimates are obtained from regressing the employment rate in 2011 on our preferred forced migration intensity measure, which uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates from both 1941 and 2011. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

empirically by examining whether the forced migrations lead to short and long-term changes in local labour market composition. We focus on the shares of local workers employed in different sectors (agriculture, manufacturing, and trade) along with the share of workers per township area (labour force density).[43] To analyse short-term effects, we make use of Census data from 1941 and 1949, and estimate our regression model from Equation (2).[44] The results are summarised in Table 1.4.

Our findings in Table 1.4 suggest that the forced migration had a large positive effect on the share of agricultural workers in the local labour force. This effect ranges between a 17.1 to 18.6 percentage point increase in the local share of agricultural workers corresponding to a unit increase in forced migration intensity. The effects on labour shares in other sectors (manufacturing and trade) are mostly close to zero, while local labour force density is affected negatively, although the point estimates are not significant in the spec-

[43]The 1941 and 1949 Censuses provide information on different categories of labour shares, but data on agriculture, manufacturing, and trade shares are provided consistently in both Censuses. Nonetheless, these categories do not cover the entire labour share for each township, as a large share of workers belong to other categories (transport, construction, or 'other sectors') that are not reported consistently.

[44]The Census data was parsed from scanned PDF documents that are available online through the Hungarian Cultural Heritage Portal (https://hungaricana.hu/en/.). Matching township names across the two Census data sets leads to a smaller sample size of towns due to certain towns not being included in the 1949 Census and changes in township names.

ifications that include county fixed effects. The null effects on manufacturing and trade, taken together with the positive effect on agricultural shares, imply that labour shares may have declined in other sectors, such as construction or transport, that we do not have consistent data on. It is also possible that small (and statistically insignificant) declines in the labour shares of all other sectors contributed to the increase in agricultural shares.

Table 1.4. OLS Results - Labour Shares in 1949

	Manufacturing Share of Labour			Agriculture Share of Labour		
	(1)	(2)	(3)	(4)	(5)	(6)
Forced Migration Intensity (Registry Data)	0.001	0.001	0.003	0.186***	0.171**	0.179*
	(0.074)	(0.072)	(0.083)	(0.070)	(0.073)	(0.095)
Number of Townships	833	664	664	833	664	664
R^2	0.000	0.254	0.304	0.016	0.116	0.185
Mean DV	0.18	0.17	0.17	0.11	0.11	0.11
SD DV	0.62	0.61	0.61	0.19	0.20	0.20
Panel B:						
	Trade Share of Labour			Labour Force Density		
	(1)	(2)	(3)	(4)	(5)	(6)
Forced Migration Intensity (Registry Data)	-0.001	0.008	0.000	-0.157*	-0.069*	-0.040
	(0.011)	(0.012)	(0.012)	(0.092)	(0.039)	(0.046)
Number of Townships	833	664	664	833	664	664
R^2	0.000	0.170	0.207	0.000	0.474	0.500
Mean DV	0.02	0.02	0.02	0.37	0.31	0.31
SD DV	0.07	0.05	0.07	1.19	0.20	0.20
Covariates	No	Yes	Yes	No	Yes	Yes
County FE	No	No	Yes	No	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 1949) on our preferred forced migration intensity measure, which uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects and all our covariates. Covariates include the 1941 labour market shares of different sectors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

We can also assess the long-term effects of the forced migrations on local labour shares. The most recent data on labour shares and labour force density is from the 2011 Census. We once again estimate Equation (2) using outcome data from 2011. In the most demanding specification, we add local district area (*'jaras'*) fixed effects, which allow us to control for local determinants of labour shares at a highly granular regional level.[45] Our results are summarised in Table 1.5.

Our findings in Table 1.5 suggest that the forced migrations can be associated with a persistent long-term effect on agricultural labour shares in each township. These effects are smaller than the short-term effects observed prior, and range between a 3.1 to 6.8 percentage point increase in response to a unit increase in forced migration intensity. The size of these effects suggests that while the initial effects (see Table 1.4) on agricultural shares persisted to the present, they did get smaller over time. It is also worth noting that these effects are no longer significant once we account for area fixed effects, which suggests little (longer-term) variation in the effect across townships within small geographic areas. [46] Our results also indicate small positive and small negative effects for manufacturing and trade, respectively, but these are only significant in some specifications. Nonetheless, the positive effect on manufacturing is robust across the specifications that include area fixed effects. A possible explanation for this is that while townships affected by forced migration moved towards agriculture initially, investments in local manufacturing activities were scaled up in the last few decades to provide a boost to local economic activity, which was still lagging behind other parts of the country during the 1980s and 90s (see Table 1.1). Finally, the effect on labour force density continues to be negative, but is not significant in most specifications.

Our findings provide clear evidence of lasting changes in local labour market composition in response to the forced migrations. These changes may have influenced divergence in economic activity across townships. A plausible story to explain this – one that is partly based on historical and anecdotal evidence (see Section 1.2) – could be summarised the following way. The displacement of the German minority population, along

[45] We have information on the current (2011) local district area designations for each town in our sample.

[46] Since the average area ('jaras') only contains 19 townships, it is possible that this is simply due to limited variation in longer term outcomes at this highly granular regional level.

with the efforts to reallocate formerly German-owned land to 'native' settlers, led to significant skills mismatches in local labour markets, and a subsequent increase in activities (mostly agricultural) that the new, relatively unskilled, labour force could partake in. This mismatch forced affected townships – which were similar to other townships in their share of agricultural labour prior to the expulsions (see Figure 1.2) – to employ a higher share of their workforce in agricultural activities, while the rest of the economy increasingly shifted towards other, more productive, sectors.[47]

[47]The post-migration period in our sample corresponds to a period of Communist rule and strong central planning policies in Hungary. At the township-level, cooperatives ('termeloszovetkezet') became the primary unit of economic activity. Cooperatives required members (joining cooperatives was state-mandated) to share all production inputs and outputs, all forming the property of the collectivist state. Under this system, it is unclear how much discretion township areas had over economic decisions that affected their workforce, although it is likely that local changes in skills composition would have influenced centrally planned policies as well. After 1968, there was a shift towards more market-oriented economic policies introduced through the New Economic Mechanism (see Balassa, 1983).

Table 1.5. OLS Results - Labour Share in 2011

Panel A:	Manufacturing Share of Labour			Agriculture Share of Labour		
	(1)	(2)	(3)	(4)	(5)	(6)
Forced Migration Intensity (Registry Data)	0.021	0.016	0.037*	0.057**	0.068***	0.031
	(0.015)	(0.015)	(0.019)	(0.023)	(0.024)	(0.020)
Number of Townships	1144	1144	1128	1247	1247	1232
R^2	0.216	0.222	0.354	0.217	0.246	0.467
Panel B:	Trade Share of Labour			Labour Force Density		
	(1)	(2)	(3)	(4)	(5)	(6)
Forced Migration Intensity (Registry Data)	-0.017*	-0.019**	-0.016	-0.083*	-0.075	-0.021
	(0.009)	(0.009)	(0.010)	(0.047)	(0.050)	(0.038)
Number of Townships	1194	1194	1179	1270	1270	1255
R^2	0.094	0.103	0.314	0.678	0.683	0.782
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Labour Shares in 1941	No	Yes	Yes	No	Yes	Yes
Area FE	No	No	Yes	No	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 2011) on our preferred forced migration intensity measure, which uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates. Covariates include 1941 labour market shares of different sectors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

1.4.2.3 Persistent Effects on the Agriculture Sector in Local Labour Markets

Simply put, our results from the previous section indicate that the share of agricultural labour increased, over time, in high forced migration towns relative to low (or no) forced migration towns. We can clearly observe this effect when looking at the two-period means of agricultural shares in Figure 1.6. High forced migration townships started at a lower agricultural labour share in 1941, when compared to unaffected townships, and ended up on a higher one by 2011, while the overall (country-wide) trend in these shares is a downward one.

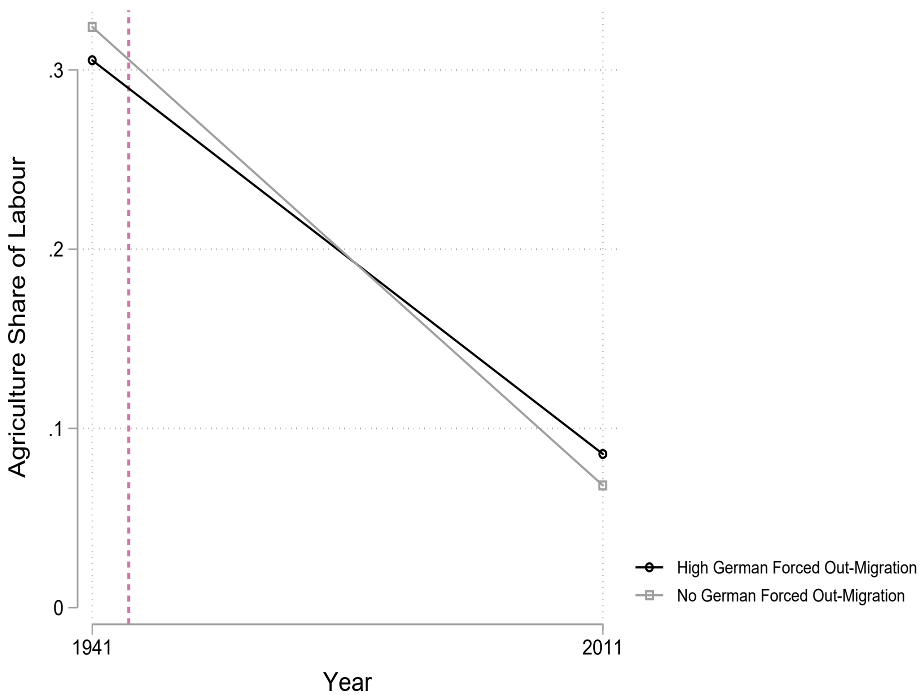
Here, we examine whether forced migrations affected the local composition of the agricultural sector. We collect data from the 2007 Hungarian Agrarian Census on three outcomes: 1) the share of land used from total agricultural land available 2) the number of agricultural producers per total agricultural area available and 3) the share of agricultural firms per township population. We look at these outcomes to assess the extent to which local agricultural supply and productivity is in line with the higher relative share of agricultural workers in high forced migration townships. We estimate the same set of specifications as in the previous sections. The results are summarised in Table 1.6.

The positive effect observed on the share of agricultural land used in Table 1.6, while only marginally significant, implies that high forced migration intensity is correlated with increased agricultural activity in affected townships on the long run. Nonetheless, the results for the share of agricultural producers and firms – which are negative and insignificant for the earlier, and positive, but very close to zero for the latter – suggest that this does not materialise in a persistent long-term influence on the presence of agricultural producers. It is possible that this is because agricultural production in these towns is characterised by large firms with substantial market power, or alternatively, it simply implies low productivity in these areas for agricultural activities.

We also provide suggestive evidence on agricultural productivity in townships affected by forced migration, by assessing the historical correlation between forced migration intensity and current crop yields. To do this, we estimate Equation (2) using crop yield data (from 2010) on 387 Hungarian agricultural areas from the FAO GAEZ database. We match each agricultural area to the nearest township to obtain values for the explanatory variable (forced migration intensity). Note, that we have no information on pre-migration crop yields, and therefore our estimates are likely only able to uncover historical correlations. Nonetheless, we include county and district area fixed effects to keep spatial determinants of crop yield relatively similar across the townships that we compare. The results are summarised in Table 1.7. We distinguish between capital and labour-intensive crop types to see if the forced migrations, and subsequent changes in local labour market composition, are associated with a specialisation in different types of crops. Our results indicate a negative association for almost all types of crops (along with total crop yield), suggesting that agricultural productivity is currently lower in high forced migration townships, although these estimates are imprecise and not significant at any reasonable level.

While the evidence presented in this section is only suggestive, it seems likely that despite the fact that the forced migrations had a long-term positive effect on agricultural labour shares in affected townships, this did not go hand in hand with increased agricultural productivity. Nonetheless, whether this is due to the forced migrations setting affected townships on inefficient economic trajectories or some other, unexplained, factor affecting long-term development, is unclear.

Figure 1.6. Share of Agricultural Labour Over Time - High vs No Forced Migration Towns



Notes: The measure of forced migration intensity here is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. The black connected line with the circles plots the change in share of agricultural labour from 1941 to 2011 for high (top 10th percentile) forced migration townships. The grey connected line with the squares shows the same outcome for townships where no forced migration had taken place.

Table 1.6. OLS Results - Agriculture Outcomes in 2007

	Share of Agricultural Land Used			Agricultural Producers per Area			Agricultural Firm Share		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forced Migration Intensity (Registry Data)	0.014*	0.018*	0.018	-0.113	-0.070	-0.176	0.004*	0.003	0.004
	(0.008)	(0.009)	(0.011)	(0.121)	(0.135)	(0.185)	(0.002)	(0.003)	(0.002)
Number of Townships	1260	1260	1245	1260	1260	1245	1270	1270	1255
R^2	0.092	0.121	0.208	0.076	0.130	0.208	0.060	0.089	0.201
Covariates (1941)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Additional Controls	No	No	Yes	No	No	Yes	No	No	Yes
Area FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 2007) on our three forced migration intensity measure. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table 1.7. OLS Results - Crop Yield in 2010

	Crop Yield (mtu/h) - Total	Crop Yield (mtu/h) - Capital Intensive Crops				
	(1) All Crops	(2) Wheat	(3) Maize	(4) Barley	(5) Simolina	(6) Other Cereals
Forced Migration Intensity (Registry Data)	-6.267	-0.218	-0.658	-0.249	0.278	-0.115
	(12.529)	(0.402)	(0.559)	(0.328)	(0.228)	(0.442)
Number of Areas	387	387	387	387	387	387
R^2	0.585	0.657	0.678	0.654	0.541	0.760
	Crop Yield (mtu/h) - Top Crops	Crop Yield (mtu/h) - Labour Intensive Crops				
	(1) Top 5 Crops	(2) Potatoes	(3) Beans	(4) Sugarbeets	(5) Fruits	(6) Vegetables
Forced Migration Intensity (Registry Data)	-6.779	-0.849	-0.042	-4.021	-0.184	-1.136
	(9.765)	(2.383)	(0.127)	(5.798)	(0.752)	(2.417)
Number of Areas	387	387	387	387	387	387
R^2	0.606	0.719	0.578	0.540	0.615	0.711
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Covariates (1941)	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The point estimates are obtained from regressing yield data on various crops (measured in 2010) on our Registry based forced migration intensity measures. Crop yields are measured in metric ton units per hectare. Crop yield data was accessed using the FAO GAEZ database. All specifications include county fixed effects, area ('jaras') fixed effects, and a full set of township level covariates from both 1941 and 2011. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

1.4.2.4 Settler Inflows

Our baseline analysis examines the reduced form effects of forced migrations but these overlap with (at least) one other treatment effect: the replacement of German households with native settlers and refugees from surrounding countries. It could be the case that there is heterogeneity in the effects of forced migration based on the settler population that replaced Germans. In this section we provide suggestive evidence on these effects of settlers.

Historical sources suggests that the land and properties formerly owned by German families were mainly redistributed to two groups (see Section 1.2):

1. Native settlers from major agricultural regions of Hungary, to whom land was allocated through postwar land reforms and
2. Refugees from surrounding countries, mainly Czechoslovakia (Slovakian border territories) and Yugoslavia (the border territory of Vojvodina).

We use data on county-level estimates from 1946-47 of native settler and Slovakian refugee groups and create dummy variables taking the value of one when native or refugee settlers migrated to a particular county in larger numbers than the inflow of migrants from each group in the median county. We then interact these dummy variables with our forced migration intensity measure and report the marginal effects by settler/refugee group. Note that these measures are likely highly endogenous to forced migration intensity itself, as settler and refugee populations moved into the properties confiscated from German families. The inflows of native settlers might be particularly endogenous if, due to preferential treatment by authorities (see Section 1.2), they had more discretion over migration location choices compared to refugee populations, and ended up in the more 'attractive' townships that Germans were expelled from. Nonetheless, there is still some regional variation in destinations across different settler groups, partly explained by the fact that it was more convenient to house refugees nearer to their country of origin, and therefore refugees mostly ended up in a few border counties (Toth, 1993). Also, while inflows of Hungarian settlers into each county are positively correlated with forced migration intensity, there are outlier counties where many townships were subjected to forced migrations with little to no settler inflow to replace the German families who were deported (see Figure A.3). We examine how marginal effects of forced migration intensity on economic outcomes correlate with these regional differences in allocation as evidence potentially suggestive of heterogenous effects across settler/refugee populations. Our results are reported in Table 1.8. The results summarised in Table 1.8 indicate negative marginal effects on population density for high native settler inflow areas, and positive marginal effects for high refugee inflow areas, implying that the overall negative effect on population density is possibly driven by areas that native settlers moved into in large numbers.

We also report marginal effects by high versus low Hungarian settler inflow categories for our labour share outcomes from both 1949 and 2011.[48] The results are reported in Figure 1.7. Overall, our results suggest that high Hungarian settler inflows potentially exacerbated the negative effects on population density, and had a positive marginal effect on the agriculture share of labour, at least on the long run.[49] While these evidence are only suggestive, they do imply that the baseline effects we observe are possibly driven by areas that Hungarian settlers moved into in large numbers. This would be largely consistent with historical sources which suggest that native settlers were largely moving from agricultural areas but were lacking the skills to fit their new local labour markets, leading to adverse effects (Toth, 1993). Historical evidence also suggests that native settlers were in some cases only interested in the removal of tangible goods and

[48]Note, that the 1949 sample is already quite small and therefore when we split effects into subcategories our estimates lose precision.

[49]It is possible that we are simply unable to detect a short-term effect in Panel A of Figure 1.7 due to small sample size in the (split) 1949 sample.

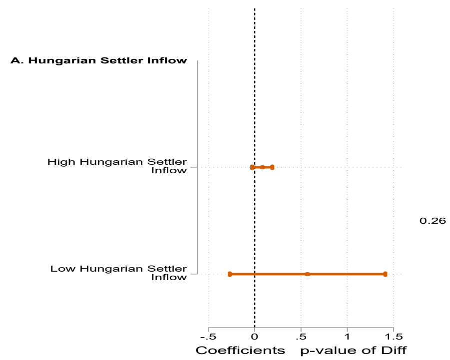
physical capital, leading to lower population density both through urban decay and through 'settlers' fleeing townships after looting them (Toth, 1993, Marchut, 2014). As explained earlier, the resulting destruction of physical capital may also partly explain the increasing agricultural focus of these townships in the years and decades after. So despite the initial mechanical increase in population density from native settler inflows, the wider negative impacts on the development of townships resulted in comparatively smaller population density overall.

Table 1.8. OLS Results - Population Density Over Time - Settler Inflows

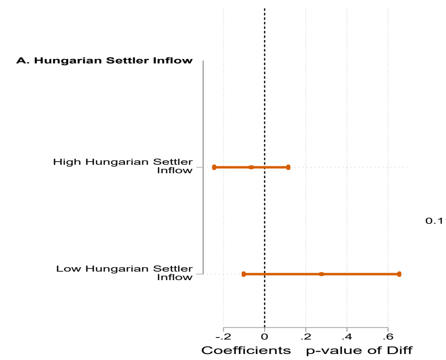
	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Forced Migration Intensity x Post x High Hungarian Settler Inflow	-0.304***	-0.357***	-0.286***	-0.263***	-0.213***
	(0.068)	(0.058)	(0.053)	(0.052)	(0.049)
Forced Migration Intensity x Post x High Slovakian Refugee Inflow	0.088	0.143**	0.131**	0.138**	0.126**
	(0.073)	(0.065)	(0.057)	(0.054)	(0.050)
Observations	8770	11614	14505	17380	20236
R^2	0.970	0.947	0.933	0.935	0.942
Town FE	Yes	Yes	Yes	Yes	Yes
County x Year FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: The point estimates are obtained from regressing the log of population density on our preferred forced migration intensity measure interacted with dummy variables indicating high settler/refugee inflows at the county-level. The interaction terms are 1) a dummy indicating higher than median numbers of Hungarian settlers in 1946, and 2) a dummy for higher than median numbers of Slovakian refugee inflows in 1947. All specifications include town fixed effects, county times year fixed effects, and year fixed effects. Observations are the number of township-years. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

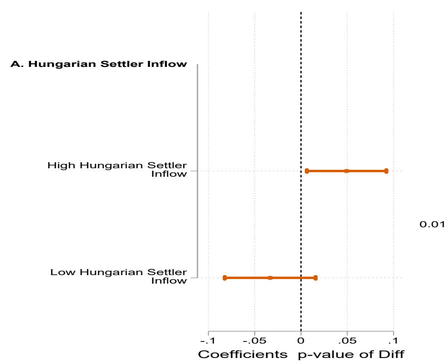
Figure 1.7. Heterogeneity by Settler Inflow



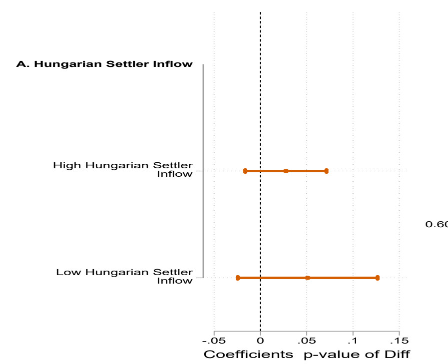
(a) Agriculture Share of Labour (1949)



(b) Manufacturing Share of Labour (1949)



(c) Agriculture Share of Labour (2011)



(d) Manufacturing Share of Labour (2011)

Notes: This figure presents heterogeneous effects of different subgroups of counties based on their level of Hungarian settler inflows in 1946-47. We interact the forced migration intensity variable with an indicator for high/low settler inflows. Confidence intervals (horizontal bars) not spanning zero indicate significance at the 5% level.

1.4.2.5 Alternative Channels for Persistent Economic Effects

The previous sections provided evidence of short and longer-term changes in local economies associated with the forced migration of Germans after WW2. Nonetheless, German families had a strong influence on their local communities, and it is likely that both their expulsion and their legacy have shaped the social, cultural, and economic development of the townships they resided in. These wider impacts from the expulsions, and the legacy of German minority presence could therefore continue to shape present-day economic outcomes through a variety of channels. In this section, we investigate a number of these potential channels. Since we do not have information on pre-treatment outcomes, we uncover historical correlations with township-level forced migration intensity.

Trust. We first examine whether the forced migrations are associated with lasting differences in local trust levels. There is strong evidence in the economics literature that historical events can have permanent effects on the trust levels of different communities (see Nunn and Wantchekon, 2011; Becker et al., 2016). In the Hungarian context, historical and anecdotal evidence strongly suggests that the expulsion of Germans had a negative influence on the social fabric of affected townships: in the decades after the expulsions, conflicts between long-time residents and settlers were common over issues such as land allocation, the perceived work effort of settlers, and political differences (Toth, 1993; Marchut, 2014). To examine whether forced migration intensity is associated with present-day levels of trust for residents of Hungarian townships, we use data from the LiTS survey to measure various forms of trust in 2016 (see Section 1.3). The results for our regression specification using Equation (2) are summarised in Table 1.9. Unlike in the previous specifications, these regressions estimate the effects of (township-level) forced migration intensity on individual-level outcomes. We control for various observable individual (age, gender, education, religion, etc.) and family (parents' education, region of birth, etc.) characteristics that could influence trust levels by making use of the rich socio-demographic information available in the LiTS data. We also control for township-level information on economic outcomes from 1941, and control for counties affected by Ottoman occupation centuries ago to account for the persistent cultural influence of this period.[50]

The results summarised in Table 1.9 indicate large (and robust) negative associations between forced migration intensity and the extent to which residents trust their neighbours, and the extent to which they trust the country's legal system (courts). For example, point estimates suggest that a unit increase in forced migration intensity can be associated with a 16.2 to 16.8 percentage point reduction in the likelihood of trusting one's neighbours. These findings suggest that lower trust in neighbours, and in courts, are correlated with a higher township-level exposure to the expulsion of German households. Note, that since we cannot control for the pre-migration levels of trust, these results should be interpreted as historical correlations. Whether the forced migrations, or pre-existing differences in trust levels across German and non-German townships, are at the root of these correlations, is unclear. Nonetheless, the idea that the forced migrations are behind these effects is strongly supported by historical evidence, which explicitly refers to increased social conflict at the local level after the expulsions (Marchut, 2014, Seewann, 2012).

Human Capital Accumulation. In this section, we examine whether the forced migrations had a lasting impact on human capital accumulation at the township level. If the settler populations were on average less skilled or educated compared to the German

[50] Note, that county fixed effects are not included in our main specification as the LiTS data only provides information on between 1-3 townships per county, leading to very small clusters of observations. Our coefficient estimates do not change their sign when these fixed effects are included, the estimates however become less precise due to inflated standard errors.

Table 1.9. OLS Results - Trust in the Present Day

Panel A:	Trust in Government		Trust in Courts		Trust in Neighbours	
	(1)	(2)	(3)	(4)	(5)	(6)
Forced Migration Intensity (Registry Data)	0.029	-0.138	-0.135*	-0.348**	-0.168***	-0.162*
	(0.071)	(0.145)	(0.078)	(0.150)	(0.044)	(0.087)
Female	0.010	0.015	0.020	0.024	0.008	0.017
	(0.020)	(0.022)	(0.032)	(0.036)	(0.028)	(0.027)
Age	0.001	0.000	0.001	0.001	0.003***	0.003***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnic Minority	0.053	0.044	-0.028	-0.094**	0.023	0.007
	(0.083)	(0.066)	(0.072)	(0.043)	(0.112)	(0.128)
Observations	922	858	922	858	922	858
R^2	0.012	0.039	0.021	0.065	0.037	0.049
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Town-level Covariates (1941)	No	Yes	No	Yes	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 2016) on our preferred measure of forced migration intensity, which uses the population adjusted measure relying on Registry data. All trust measures are dummy variables where a value of one indicates some level of trust in the specific entity. All specifications include control variables for individual and family characteristics, along with 1941 levels of town-level covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

population they replaced (Marchut, 2014), this could have had a permanent effect on the composition of human capital and educational preferences in affected townships, which could then become important factors in their long-term economic development. To test this, we estimate Equation (2) using educational measures – the share of residents with no schooling, the share with a high school degree, and the share with a higher degree – from the 2011 Census as our outcomes of interest. Our results are summarised in Table 1.10. The point estimates presented in Table 1.10 indicate that forced migrations had no (long-term) effect on the share of residents with share of residents with a higher (university or vocational college) degree.[51] On the other hand, we do observe a negative effect for the share of locals with no schooling, although estimates are not significant in all specifications, and point estimates are very close to zero. Overall, our results do not

[51]Note, that schooling in Hungary is compulsory until age 16.

suggest that changes in human capital accumulation are behind the persistent effects we observe, although we cannot rule out that such effects were present in the short and medium-run after the expulsions.

Table 1.10. OLS Results - Human Capital in 2011

	Share with No Schooling			Share with High School Degree			Share with Higher Degree		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forced Migration Intensity (Registry Data)	-0.008***	-0.006**	-0.001	0.009	-0.016	-0.012	0.013	0.002	0.005
	(0.002)	(0.003)	(0.003)	(0.018)	(0.022)	(0.021)	(0.010)	(0.011)	(0.008)
Number of Townships	1270	1270	1255	1270	1270	1255	1270	1270	1255
R^2	0.048	0.120	0.228	0.383	0.642	0.726	0.345	0.506	0.654
Covariates (1941)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Area FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 2011) on our preferred measure of forced migration, which is the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates, along with the shares of people in occupations in 1941. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Religious Composition Finally, we test whether the forced migrations had a persistent impact on the religious composition of the population at the township level. Religious composition can play an important role in influencing historical economic development at the regional level (see Becker et al., 2021). The Hungarian German minorities were overwhelmingly affiliated with the Roman Catholic Church (Marchut, 2014; Grosz, 2020), albeit in this respect they were not different from the rest of the country, where Roman Catholicism was the majority religion in most counties before WW2 (see Figure A.4). According to Marchut (2014), there were very few Protestant German communities, and the share of Protestants among the German minorities were lower than the national share. Moreover, it is unclear what the religious affiliations of the settlers were in each township, although we can make generalisations based on certain subpopulations. For example, historical sources reveal that some of the settler populations, such as the Szekelys of Bukovina or the Csango's, were also predominantly Catholic

(Gatti, 2019). Based on the information available, it is therefore unclear whether the population changes led to changes in religious composition.

We test whether the forced migrations are associated with persistent changes in township-level religious composition by estimating Equation (2) using 2011 Census data on the share of Catholics, Protestants, and non-religious residents in each township. We use county-level religious shares from 1941 in some specifications, and county and area fixed effects in others to control for the possibility that our inference is biased by the presence of spatial autocorrelation of religious composition. The results are presented in Table 1.11. Our results suggest a non-robust negative association with Protestant shares, and a positive association with Catholic shares, which are no longer significant once we include county fixed effects. The effect on non-religious shares is close to zero and not significant, and overall there is little evidence that the expulsion of the German minority population had a lasting impact on the religious composition of affected townships. Based on these findings, and based on the fact that on average, most settler sub-populations (for whom we have this information) were likely to share the same religion with the Germans they replaced, we conjecture that it is unlikely that changes in religious composition are behind the baseline results we observe.

Table 1.11. OLS Results - Religious Composition in 2011

	Share of Catholics			Share of Protestants			Share of Non-Religious		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forced Migration Intensity (Registry Data)	0.096***	0.003	-0.009	-0.059***	-0.019	-0.025	-0.024	0.002	0.016
	(0.034)	(0.032)	(0.031)	(0.020)	(0.021)	(0.017)	(0.018)	(0.018)	(0.019)
Number of Townships	1270	1270	1255	1270	1270	1255	1270	1270	1255
R^2	0.311	0.348	0.534	0.376	0.390	0.580	0.441	0.459	0.617
Covariates (1941)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Area FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The point estimates are obtained from regressing our outcomes (measured in 2011) on our preferred measure of forced migration, which is the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

1.4.3 Threats to Identification

Measuring forced migration intensity. The Registry data we use to measure forced migration intensity is generally considered to be the best source of data on the forced migrations, but is still likely to contain errors due to mistakes by authorities (Marchut, 2014). We therefore use two additional measures of forced migration intensity as robustness checks to examine the sensitivity of our findings to alternative measures of forced migration intensity. First, we use the pre-migration share of Germans in Hungarian townships, measured in 1941, as a proxy for township-level exposure to the expulsions. Our use of this measure follows similar studies such as Acemoglu et al. (2011), Arbatli and Gokmen (2018), Clemens et al. (2018b), and its advantage is that *ex ante* information on the shares of Germans is plausibly exogenous in the sense that it is unaffected by potential responses to the expulsions. We also use a measure of forced migration intensity calculated using the formula described in Section 1.4, but based on the number of Germans in each township from the 1949 Census instead of the data from deportation registries. Results using these measures for our main outcomes are summarised in the Appendix (Table A.7 to Table A.9, along with Figure A.5). Overall, our results are robust across different measures of forced migration intensity.

Matching specification. In our baseline specification in Section 1.4, we control for a wide range of geographic and demographic factors that could drive changes in township-level economic development. We also include town, time, and county time fixed effects, meaning that most of the effects we identified in the previous sections controlled for pre-existing differences in outcomes between high and low forced migration townships, and were based on within-region variation where we can expect relatively low variance of key geographical factors. Nonetheless, it is still possible that some unobservable factor has influenced the initial share of Germans in different townships – leading to higher forced migration intensity for these localities – while at the same time having a persistent influence on economic development. For example, the initial German settlements were mostly in areas ravaged by Ottoman occupation in the centuries prior (see Section 1.2), which were less developed than other parts of the country. While we control for fixed regional factors that are likely to influence these outcomes, it is not implausible that specific local geographic factors could have simultaneously influenced Ottoman presence, wartime losses in terms of population, and long-term economic outcomes. To check the sensitivity of our findings to this possibility, we employ a propensity score matching (PSM) approach to create matched pairs of townships with high and low forced migration intensity that are similar in exogenous geographic characteristics and also in endogenous factors that could drive German minority presence.[52] High and low forced migration intensity towns are matched based on their pre-migration economic and demographic characteristics (local labour market and demographic composition) as well as geographic variables (arable and non-arable land area, area of land suitable for cultivation, etc.).[53] Note, that while creating matched pairs this way arguably leads to a comparison of towns more similar in their observable characteristics, it also leads to reduced statistical power through a smaller sample size.[54] We re-estimate the specifications from Section 1.4.1 using the samples of matched pairs we created. The results are summarised in the Appendix, Table A.10 to Table A.13. Our main findings from the last two sections are all robust to using the matched sample specifications.

Treatment effect heterogeneity. Our main specifications follow a multiple time period (when population density is the outcome) or two-period (for other outcomes)

[52]Since our specification where trust levels are the outcome rely on individual survey data matched with township level forced migration data, we cannot employ this approach to the specification summarised in Table 1.9.

[53]For our analysis of the effect on population density, where we have longitudinal panel data available for the period 1920 to 1990, we match based on time-variant township characteristics.

[54]When matching towns, we impose a caliper of 0.01 so that we only keep 'good' matches based on observable baseline characteristics.

continuous treatment difference-in-differences approach. This approach was initially outlined in Card (1992), and relies on the fact that treated units (townships) are each affected with a different dose 'd' of the treatment (forced migration). As recently outlined in Callaway et al. (2021), researchers using continuous difference-in-differences can estimate a range of causal parameters. First, we can identify the average level effect (ATT) for any non-zero level of forced migration intensity – for any treatment dose 'd' – as the effect for the group that received dose 'd' when compared to all other groups that received some other (or no) dose. Our estimation of treatment parameters can therefore rely on multiple different counterfactual scenarios. For example, we could ask what the level effect of forced migration was for townships that experienced forced migration intensity equal to 0.1 compared to those that experienced no forced migration. We could also ask what the same effect is when comparing townships with a forced migration 'dose' of 0.1 to townships with a dose of 0.05 – in this case, the lower intensity units become the counterfactual for the higher intensity ones. Moreover, as we attempt in this paper, we can identify the average slope effect (ACR, or average causal response), which is the causal effect of moving from dose d to some other dose d' – this entails identifying the average difference between potential outcomes under some level of forced migration intensity (dose 'd') compared to potential outcomes under an incremental (marginal) change in forced migration intensity for the townships who experience dose 'd'.

The key insight from Callaway et al. (2021) for our application is that when using the two-way fixed effects (TWFE) regression specification outlined in Section 1.4, the standard difference-in-differences assumption of parallel trends is not sufficient to allow for a causal interpretation of estimated coefficients. This is because the TWFE regression weights together outcome changes that are compared across incrementally different treatment doses. Even when parallel trends hold, this can introduce selection bias in the estimates if the average level effect (ATT) is different across different treatment doses. On the other hand, TWFE estimates can still be interpreted as the ACR across all doses if the ’strong’ parallel assumption holds – in our case, this assumption requires that for all ’doses’ of forced migration intensity, the average change in outcomes over time across all townships if they received dose d, is the same as the average change in outcomes for all townships that experienced dose ‘d’ of forced migration intensity. So, for example, if all towns received a dose of 0.05, their change in average outcomes overtime must be the same as the townships that did experience a dose of 0.05 in reality. Similarly, if all towns received a dose of 0.1, their change in average outcomes overtime must be the same as the townships that did experience a dose of 0.1. This must be the case for all doses. In other words, on average across all doses, there should not be selection into a particular dose, or into a particular level of forced migration intensity. How likely is this assumption going to hold in our case?[55] In Section 1.4, we show strong evidence that most variation (see Figure 1.3 and Figure 1.4) in our treatment measures are explained by the, arguably exogenous, pre-migration presence of German minorities in Hungarian townships, and this is even true when only comparing townships with non-zero levels of forced migration. This suggests that there are unlikely to be fundamental differences between townships with different doses, and selection bias in dose levels is unlikely to be an issue given that no observable factor (other than the pre-treatment share of Germans) has been found to explain forced migration intensity. However, due to data limitations, we must maintain some caution, as we cannot test a wider range of factors that could be correlated with treatment and outcomes. Where for example, there could be some unobservable characteristic that makes deportation less likely while improving township level development. Furthermore, a limitation of our approach is that we cannot formally establish the validity of the strong parallel trends assumption in the context of this study.

[55]The latest version of the Callaway et al. (2021) study does not yet provide formal ways to validate the stronger parallel trends assumption.

1.5 Conclusions

Forced migrations can not only affect receiving populations and alter the life experiences of the migrants themselves, but can also lead to permanent changes in sending economies. In this paper, we used historical and contemporary data to examine the economic and social effects of the postwar expulsions of German minorities on Hungarian townships. We found that forced migrations had a lasting negative impact on local economic development and trust levels in affected townships, and led to fundamental changes in the skills composition and industry specialisation of local economies. More specifically, our results indicated that townships affected by forced migrations increasingly specialised in agricultural activities in subsequent decades, and yet we could not find evidence that this has led to improved agricultural productivity in these areas.

While we consider a number of possible mechanisms and explanations for the empirical patterns uncovered in our analysis, a few questions remain. First, the lack of high-quality data on economic activity and productivity at the township level makes it difficult for us to distinguish between population changes and changes in local labour markets as potential drivers of long-term regional divergence in economic development. As we rely on measures such as population and labour force density to assess changes in economic activity over time, it is unclear whether these changes are a result of a failure to replace the German populations and labour force, or are due to structural differences between German and settler populations. Second, in this study we were unable to examine other potential factors behind persistent changes in local economic development that may have been induced by the forced migrations. For instance, it is still unclear how (and why) the effects of forced migrations differ across townships with different types of settler populations, and more granular data would be needed to assess the precise effect of the expulsions on the skills composition of the local labour force. Moreover, the period we are studying in this paper is partially overlapping with the period of strong

Communist central planning policies in Hungary, and how these policies affected the townships that were exposed to the forced migrations is as of yet poorly understood. Future studies should investigate these questions in more detail.

Nonetheless, our study offers important lessons on the issue of forced migration, which continues to affect millions of displaced individuals worldwide. The key takeaway is that forced migration can have negative effects on origin economies, even (or especially) in cases where it is motivated by redistributive objectives. In other words, even those areas that are specifically meant to benefit from forced migration may end up worse off in terms their economic development and social fabric. Considering the traumatic and life-altering experiences of those forcibly removed from their homes, this insight suggests that forced migration is not only ethically and morally reprehensible, but is also not a sensible policy from an economic and social perspective.

Chapter 2

Migration Trends in Scotland and Why Measurement Matters

2.1 Introduction

The impact immigration can have on native wages is an important topic at the forefront of the migration policy debate in many countries worldwide. However, across many studies there appears to be conflicting result, and it is important to understand why this is the case, to better inform policymaking. In recent years there has been an advancement in understanding why immigration studies find different results that can be applied when reviewing the literature. Dustmann et al. (2016) show that it can vary depending on the empirical specification used, where they categorise studies three types of specification: Skill-Cell, Spatial and Mixed Approaches, each of which measure a different relationship between immigration and native wages. While Card and Peri (2016) show that how you measure migration is also important, where if we define immigration as the share of the local labour force, that I call the Immigrant Share, then estimates will be negatively biased as a result of relative-local demand shocks increasing native inflows and wages. They then provide an alternative measurement, the Migrant Inflow, that they show is not negatively biased. This paper extends this analysis to a third variable commonly used in the UK migration literature, the Migrant Native Ratio.

Overall, I discuss three measurements of immigration: the Migrant Inflow, Immigrant Share, and Migrant Native Ratio. Extending the findings of Card and Peri (2016) I show due to a positive correlation between native inflows and native wages then in addition to the Immigrant Share measurement, the Migrant-Native Ratio could also result in a negative bias when estimating the impact of migration on native wages. I then review key literature on the impact of migration on wages, where I show that different measures of migration are favoured across studies depending on the empirical specifications used and year it was published. Finally, using UK data, I estimate the impact of immigration on native wages using the three different measurements of migration. I find that consistent with Card and Peri (2016), the Immigrant Share does empirically show a negative bias, however unexpectedly this is not the case for the Migrant-Native Ratio.

To explore the potential bias that can occur when measuring migration, I first review Card and Peri (2016)'s findings. To begin, I reproduce the first-order approximation of the Immigrant Share, and explain why a negative bias can occur if native wages and native inflows are positively correlated, for example due to relative local demand shocks. I then take the first-order approximation for the Migrant-Native Ratio, a common measurement in the UK literature, and show that this could also result in a negative bias for the same reason but has yet to be explored empirically. To estimate whether the three different measurements of immigration produce different estimates of the impact of immigration on native wages I use the 2004-2019 Annual Population Survey (APS) from the Office for National Statistics (ONS). I disaggregate UK regions into 18 Government Office Regions and 36 adjusted NUTS2 regions. For each regional definition I regress native wages on each of the three measurements of immigration separately. To address the endogeneity that can arise from immigrants moving into areas with relatively positive demand shocks, I use the standard shift-share instrument commonly used in the literature (Card, 2001), where I use past immigrant shares from the 2% Sample of Anonymised Records of the 1991 Census to construct an instrument that captures exogenous variation in migrant inflows as a result of network effects. When using spatial variation, it is possible that estimates of the wage impact of migration are underestimated as in response to migration natives may move out of their region, spreading the local supply shock to other regions (Borjas, 2006) or discouraging natives to move into areas with high levels of immigration (Dustmann et al., 2017, Ortega and Verdugo, 2022). At the same time, Card and Peri (2016) show that the negative bias from using the Immigrant Share measurement is even worse when estimating native outflows. Using the same specification I separately regress native outflows on the three measures of migration.

In line with Card and Peri (2016), I find that when estimating the impact of migration on native wages, using the Immigrant Share to measure migration results in a negatively biased point estimate, and is just under double the size of the point estimate obtained when using the unbiased Migrant Inflow measurement. Unexpectedly, the point estimate obtained using the Migrant-Native ratio was very similar to the point estimate when using the Migrant Inflow. This is in contrast to Edo (2020) who finds no difference between the Migrant Inflow and Immigrant share definition of migration when estimating French-Algerian repatriation in 1960s France. In contrast to previous UK literature (Dustmann et al., 2013, Lemos and Portes, 2013) we find the average total is negative. When using 18 government office regions we find a 1 percentage point increase in Migrant Inflow decreases native wages on average by -0.145%, while across 36 NUTS2 regions it decreases by wages -0.322%. This could potentially be explained by the shift in lower paid occupations to rely more heavily on immigration, where Dustmann et al. (2013) and Sá (2015) find that natives in the lowest wage decile were more negatively impacted by migration.

I find that it is unlikely that these results are underestimated as a result of native outflows as I find insignificant results when estimating the impact of migration on native outflows for all three measures of migration. Considering the insignificant coefficients, there is weak evidence of a negative bias when estimating the impact on native outflows. When grouping regions into 18 Government Office Regions, Ordinary Least Squares (OLS) results estimate a negative point estimate of -0.975 for the Immigrant Share and -0.360 for the Migrant-Native Ratio. This is in contrast to the Migrant Inflow measure that estimated a point estimate of 1.657, results are similar when using 36 adjusted NUTS2 regions.

This paper contributes to the literature in three ways. Firstly, I expand upon findings from Card and Peri (2016) and apply them to an additional measurement of migration commonly used in the UK literature. Secondly, I provide an alternative perspective to recent reviews of literature on the impact of migration on native wages (Dustmann et al., 2016, Edo, 2019). Where I explore how the measurement of migration differs across the three empirical approaches identified by Dustmann et al. (2016), and how this has developed since Card and Peri (2016). Lastly, I contribute to the literature identified by Dustmann et al. (2016) that uses a spatial approach to estimate the total impact of migration on native outcomes, which can be positive, insignificant or negative depending on the context (Card, 1990, Altonji and Card, 1991, Lemos and Portes, 2013, Dustmann et al., 2013, Basso and Peri, 2015, Foged and Peri, 2016, Mitaritonna et al., 2017, Jaeger et al., 2018, Peri and Yasenov, 2019, Edo, 2020, Ortega and Verdugo, 2022). More specifically, I contribute to the UK literature that estimates the impact of migration on UK-born native wages (Dustmann et al., 2005, 2013, Lemos and Portes, 2013, Sá, 2015) where the spatial literature finds an insignificant effect on average native wages, but a positive effect of those on the top of the wage distribution and a negative effects for those on the bottom.

The rest of the article proceeds as follows: Section 2.2 describes the data sources, Section 2.3 breaks down different measurements of migration and their use in the literature. In section 2.4 I discuss my estimation specification and identification strategy and in section 2.5 I discuss my estimation results. Section 2.6 concludes.

2.2 Data, Measurements and Summary Statistics

We use data from the UK Annual Population Survey (APS) from 2004-2019 to provide an analysis of how different ways to measure migration can have an impact on the results of a spatial regression model of the impact of migration on native wages.

2.2.1 Data

I use the Annual Population Survey (APS), which provides detailed data on labour outcomes and migration for a large, representative sample for the UK with boosted samples for smaller regions. The APS consists of repeated cross sections and contains year data for the years 2004-2019. The APS is a survey of private households in the UK conducted by the Office of National Statistics (ONS) in Great Britain and Northern Ireland.[1] The sample size of the APS is made up of around 320,000 households in each survey, which is the widest ranged household survey in the UK. It allows the generation of statistics for smaller UK regions, as it utilises sample boosts from the Local Labour Force Survey and APS boost in 2004 and 2005. These local boosts allow us to break down the data to regional levels while maintaining a good sample size and accuracy, where more details can be found in Appendix B.2. The APS contains data on employment, unemployment, income as well as information on age, education and occupation.

Following standard practice in the literature, I instrument country-wide migrant flows utilising the regional share of migrants from each country of origin in the 2% Sample of Anonymised Records for the 1991 Census for Great Britain and Northern Ireland, which is discussed further in Section 2.4.2. [2] [3]

[1]Office for National Statistics, Social Survey Division (2021). *Annual Population Survey, 2004-2021: Secure Access* [data collection]. 21st Edition. UK Data Service. SN: 6721, DOI:10.5255/UKDA-SN-6721-7

[2]Office for National Statistics, Census Division, University of Manchester, Cathie Marsh Centre for Census and Survey Research (2013). *Census 1991: Individual Sample of Anonymised Records for Great Britain (SARs)*[data collection]. UK Data Service 7210,DOI: 10.5255/UKDA-SN-7210-1

[3]Office for National Statistics, Census Division, University of Manchester, Cathie Marsh Centre for Census and Survey Research (2013). *Census 1991: Individual Sample of Anonymised Records for Northern Ireland (SARs)*[data collection]. UK Data Service 7212,DOI: 10.5255/UKDA-SN-7212-1

2.2.2 Measurements

I define *migrants* as those individuals interviewed by the APS that were not born in the UK. The APS records gross weekly wages and total hours worked per week. Using these two pieces of information I calculate gross hourly wages, that I deflate by the 2015 CPI. I also calculate a measurement of real hourly wage robust to outliers, where I trim observations in the 1st and 99th percentile for real hourly wages in each year.

I construct a region-year panel for the years 2004 to 2019 by aggregating wages for those who are of working age, between the ages of 16 to 64. For our regression analysis I divide the UK into 18 regions following Dustmann et al. (2013) and 36 regions according to the NUTS2 definition, where I combine Inner and Outer London into one region. Due to using the 1991 Census, when constructing the 36 NUTS2 regions, I adjust the smaller regional make-up of regions that do not align with the census data. Firstly, I ensure that the NUTS2 definition in the APS was standardised according to the NUTS2 definition in 2004. Where I moved the Unitary Authority of Halton back from the NUTS2 region of Merseyside to Cheshire in the APS data. Secondly, I ensure that the definition of regions matches the definition found in the 1991 Census data. This required three changes. Firstly, in the 1991 Census Argyll and Bute is entirely part of the Strathclyde region of Scotland, however in the NUTS2 definition in the APS it was split between the Highland and Islands and South-West Scotland. As such, I moved all those living in the unitary authority of Argyll and Bute to the South-West Scotland region which would match the definition of Strathclyde in the 1991 Census. Secondly, the 1991 Census defines two Welsh regions of Gwent and Clwyd. Gwent is made up of the unitary authorities of Caerphily, Blaneau Gwent and Torfaen. Clwyd is made up of the Unitary Authorities Flintshire and Wrexham. However, the Unitary Authorities that make up Gwent and Clwyd are split between the two NUTS2 regions for Wales: West Wales and the Valleys and, Glamorgan. To keep this standardised between the census and the APS, in the APS data I grouped the Unitary Authorities to reconstruct the region of Gwent and Clywyd. Then in both the census and the APS the former was grouped with the Glamorgan region of Wales and the latter was grouped with the West Wales and the Valleys region.

2.2.3 Summary Statistics

Table 2.1 reports selected unweighted characteristics for working age natives and migrants in the years 2004 and 2019. In both years migrants had a higher real hourly wage than natives, however this gap has narrowed from 1.25 percentage points to 0.29 percentage points. When we trim our wages by winsorising them at the 1st and 99th percentiles then this gap is only 0.11 percentage points by 2019. Across both time periods, working age migrants are slightly younger compared to natives, where it is the case that the average age of natives increased by just under two years, while migrants age maintained a similar average. The proportion of women working is also slightly higher for migrants. In terms of education, working age migrants are -on average- better educated compared to working age natives. The average educational attainment improves from 2004 to 2019, and the gap between natives and migrants remain relatively constant.

Table 2.1. Summary Statistics for Working Age Natives and Migrants in 2004 and 2019

	Natives		Migrants	
	2004 Mean	2019 Mean	2004 Mean	2019 Mean
Real Hourly Wage	13.12	14.37	14.38	14.67
Trimmed Real Hourly Wage	12.94	13.97	14.12	14.08
Age	40.91	42.27	39.55	40.08
Female	0.52	0.52	0.54	0.55
Education				
Higher(>25)	0.01	0.02	0.07	0.10
High (21-24)	0.15	0.24	0.29	0.37
Intermediate (19-21)	0.59	0.61	0.45	0.40
Low (16-18)	0.19	0.07	0.11	0.05
None/Still in Education	0.06	0.07	0.08	0.07

Notes: Entries are for working age(16-64) natives and immigrants for the average real hourl wage, trimmed real hourly wage, average age, share of female and the share in each education group in 2004 and 2019.Real hourly wages are trimmed through winsorisation, where we trim the top 99th and bottom 1st percentile of the real hourly wage distribution in each year .Higher education: left full-time education after age 25, High education: left full-time education between age 20-24, Low education: left full-time education between age 16-19, None/Still in education: left education at age 15 or below, or is still in education. Wages are trimmed at the 1st and 99th percentile in each year. N is the number of observations for each statistic. Source: APS 2004-2019

2.3 Why Measurement Matters

Since Card and Peri (2016), the literature looking at the impact of immigration on native wages has more carefully thought about how we should measure immigration. However, this has yet to be applied to the UK immigration literature. In this section I outline Card and Peri (2016)'s findings and how this can be extended to the migrant native ratio, a measurement of immigration commonly used in the UK literature. I then provide a broad overview of the measurements of immigration used in previous literature and how this varies across empirical specifications and how the preferred measurement has developed overtime.

2.3.1 Measuring Migration and Bias

In this paper I will consider three different measurements of immigration. The Immigrant Share, m_{rt}^{ML}, as shown in equation 2.1, that measures immigration as a share of the labour market for either a skill-group, region or both. Where M_{rt} is the stock of immigrants and N_{rt} is the stock of natives in labour market r, at time t. The Migrant Inflow, $\triangle m_{rt}^{MF}$, as shown in equation 2.2 that looks at the change in immigration as a proportion of the labour market from a previous time period and finally the Migrant-Native Ratio, m_{rt}^{MN}, as shown in equation 2.3 which is similar to the Immigrant Share except only the native stock is present in the denominator.

$$m_{rt}^{ML} = \frac{M_{rt}}{N_{rt} + M_{rt}} \tag{2.1}$$

$$\triangle m_{rt}^{MF} = \frac{M_{rt} - M_{rt-1}}{N_{rt-1} + M_{rt-1}} \tag{2.2}$$

$$m_{rt}^{MN} = \frac{M_{rt}}{N_{rt}} \tag{2.3}$$

Card and Peri (2016) show that if not careful when selecting an appropriate measure of migration then this can result in biased results. In their review of Borjas (2014) they take a closer look at the Immigrant Share measurement from equation 2.1. The authors show that this commonly used measure of migration can result in a negative bias when estimating the impact on native wages. To show this, Card and Peri (2016) take the first-order approximation of the Immigrant Share, that I outline in equation 2.4

$$\triangle m_{rt}^{ML} = (1 - m_{rt-1}^{ML})\frac{\triangle M_{rt}}{M_{rt-1} + N_{rt-1}} - m_{r-1t}^{ML}\frac{\triangle N_{rt}}{N_{rt-1} + M_{rt-1}} \tag{2.4}$$

As explained in Card and Peri (2016), the Immigrant Share can change due to a weighted average of two factors. Firstly, an increase in the number of migrants in the local labour market $\frac{\triangle M_{rt}}{M_{rt-1}+N_{rt-1}}$. Secondly, a decrease in the number of natives in the local labour market $\frac{\triangle N_{rt}}{N_{rt-1}+M_{rt-1}}$. When estimating the impact of migration on native labour outcomes, such as wages, this second factor could result in bias. As changes in the number of natives could be correlated with native labour outcomes in that labour market. Card and Peri (2016) explain that a positive correlation could arise if a relative-demand shock in a labour market would increase native wages and attract native workers. However, because native inflows are negatively associated with the Immigrant Share variable, then native inflows as a result of a relative-demand shock would cause a negative bias when regressing native wages on the Immigrant Share.

So, if natives, like migrants, move into groups where wage growth is high then we would expect a negative bias, which is weighted more heavily the larger last period's immigration share, m_{r-1t}^{ML}, is in a group. For example, if a group r in time t sees a 10% increase in migration which decreases native wages by 1% we would obtain a coefficient of -0.1. But if there is an endogenous native inflow of 9% then this would result in us falsely measuring a 1% change in migration resulting in a decrease in native wages and we would obtain a coefficient of -1%. They instead propose that the Migrant Inflow is a better measurement of immigration as it is unbiased as a result of the lagged denominator. They show using US data that when migration is measured by the Migrant Inflow, the results are less negative compared to the results obtained using the Immigrant share. The Migrant-Native ratio outlined in equation 2.3 is however not explored in their paper, despite it being a common measurement used in the UK literature (Dustmann et al., 2005, 2013, Nickell and Saleheen, 2017). Following Card and Peri (2016), I derived the first-order approximation for this measurement of migration shown in equation 2.5.

$$\triangle m_{rt}^{MN} = \frac{\triangle M_{rt}}{N_{rt-1}} - m_{rt-1}^{MN} \frac{\triangle N_{rt}}{N_{rt-1}} \tag{2.5}$$

This equation shows that the same negative bias as a result of native inflows could be present. Although this negative bias is similarly weighted by the previous periods migrant-native ratio, the variation due to changes in immigration is not weighted less and so we may expect a smaller negative bias.

2.3.2 Measuring Migration in the Literature

Considering that how migration is measured can potentially bias results estimating the impact of migration on native wages, then this could effect how we interpret previous studies on this topic. Dustmann et al. (2016) argued that we can split the empirical migration literature into three key types of specifications. Firstly, the National Skill-Cell approach, popularised by Borjas (2003). This approach uses nationwide variation in immigration into experience by skill cells. Where a skill group is often defined using education, but can be defined using occupations. This approach measures the partial effect of immigration on wages of high-experience natives relative to low experience natives within a skill-group. Dustmann et al. (2016) show that this approach mechnically should estimate a negative point estimate. Secondly, the Spatial approach uses variation in immigration across different spatial regions within a country. This approach measures the total effect of immigration on native wages in a region, and therefore measures a fundamentally different relationship than the skill-cell approach. The direction for the total effect is ambiguous, and so we may expect positive or negative coefficients. Lastly, the mixed approach uses changes in immigration across regions and skill-groups, where skill-groups can be defined by their education or their occupation. This approach again measures only the partial effect of immigration and identifies the relative impact of immigration on natives in high skill groups relative to low skill groups. They show that this relationship is also mechanically negative. Due to the difference in the interpretation of each of these three specifications, when comparing how different studies measure migration we must do so in the context of their specification.

Table 2.2 Panel A summarises key papers in the Skill-Cell literature, similar to Dustmann et al. (2016), with the additional information of what measure of immigration they use. These papers heavily favour the Immigrant Share measure, this is a result of them following the specification first outlined by Borjas (2003), where the results were updated in Borjas (2014). As discussed, Card and Peri (2016) replicated Borjas (2014)'s results using area-skill fixed effects instead of using first differencing finding the same negative and significant point estimate of -0.53. They then show, that if you first difference the specification and use Immigrant Share then they find a coefficient of -0.237, which becomes insignificant once they use the unbiased Migrant Inflow definition at -0.124. Since this paper only Sharpe and Bollinger (2020) use a different measurement of migration, looking at the percentage of hours worked, and they continue to find a large negative result when using occupation-experience groups.

The mixed literature, as summarised in Table 2.2 Panel B, takes a different approach to measuring immigration. As opposed to directly using the immigrant shock to a skill-region group. They instead use a measure of the supply of workers in a skill group relative to the total supply within a region, which I label as the Relative Supply. For example, Card (2001) uses the log population share of a specific occupation, region group, while Glitz (2012) uses the change in the log labour force share in a given occupation, region, time group. Typically this is then instrumented using the standard instrument outlined by Card (2001) which predicts supply driven immigration flows as a result of historical immigration shares. Glitz (2012) however takes advantage of a natural experiment in Germany, where newly arriving ethnic Germans from 1996-2001 were distributed across Germany primarily according to the proximity to family members. Borjas (2006) continues to use the Immigrant Share but this does not lead to more negative results compared to other literature using a mixed approach. Other papers instead focus on the wage gap due to changes in the relative supply of workers from two education groups, such as high-school graduates and dropouts, or college educated and high-school graduates and find an overall negative relationship as would be expected (Card and Lewis, 2007, Card, 2009, Lewis, 2011). Despite the popularity of the relative supply measure, Nickell and Saleheen (2017) who look at the variation of immigration across occupation-region groups in the UK follow Dustmann et al. (2013) and use the Migrant-Native ratio to measure immigration shocks to occupation-region groups. Despite the different measurement of immigration, it obtains a similar sized coefficient as other papers in the mixed approach literature.

The Spatial approach literature has a wider variety of approaches as shown in Table 2.3. The earlier literature initially used the Immigrant Share to measure migration (Altonji and Card, 1991, Card, 2007), although some later papers also used the Immigrant Share (Foged and Peri, 2016, Breunig et al., 2017, Mitaritonna et al., 2017). Before Card and Peri (2016), some papers used a variable similar to the Migrant Inflow variable (Boustan et al., 2010, Lemos and Portes, 2013, Sá, 2015). However, due to the variety of settings, it would be difficult to attribute any difference in findings to the choice of migration measurement.

In the UK, Sá (2015) who mainly looks at the impact of immigration on house prices, uses the 2003-2010 UK LFS to estimate the impact of Migrant Inflow on native wages along the wage distribution. While Lemos and Portes (2013) investigate the impact of immigration from EU Ascension countries from 2004-2006. However, due to the different setting these results are not directly comparable. Seminal papers studying UK migration by Dustmann et al. (2005, 2013) are unique, as they favoured using the Migrant-Native Ratio to measure migration which was later used by Nickell and Saleheen (2017) in their paper using a mixed approach. Directly comparing Dustmann et al (2013) to Sa (2015) is not possible, as they consider different time periods. However, both find that natives on the lower end of the wage distribution are more negatively effected by immigration than those in the higher end. Furthermore, Dustmann et al. (2013) provides the average effect on native wages, which is more positive than that of Lemos and Portes (2013) at 0.4 compared to an insignificant point estimate of 0.246 respectively.

Since Card and Peri (2016) had shown the negative bias present in the immigrant share variable, studies have paid more attention to how they measure migration. Basso and Peri (2015) specifically outline that they carefully chose the migration variable to avoid this negative bias. Dustmann et al. (2017) look at exogenous variation due to a policy that allowed workers from the Czech Republic to work in Germany but not live there and find a negative impact of -0.134% that they attribute to reduced demand-side effects of workers not remaining in Germany, the shock being unexpected and particularly large, and firms not adjusting their capital to the shock as they expected it to be temporary. Jaeger et al. (2018) focus on the validity of the shift-share instrument in a setting where the country of origin composition remains stable overtime. However, their paper provides results using the standard shift-share instrument using US Census data which provides a useful comparison to earlier US studies that used the Immigrant Share, although the time period and specifications differ slightly making a direct comparison not possible. The authors find a negative but insignificant point estimate of -0.193%. This is more negative than Card (2007) who find a positive and insignificant point estimate of 0.06 (0.01), but is much less negative than Altonji and Card (1991) who find a negative and insignificant coefficient of -1.1%.

Ortega and Verdugo (2022) look at the impact of immigration on native workers in France from 1976-2007 across commuter zones. In this study they use the Migrant Inflow variable, referring directly to Card and Peri (2016). Where after accounting for compositional changes to the local population as a result of immigration using a baseline commuter zone as opposed to the current commuter zone of natives, they find a negative effect on native wages of -0.238%, but only at a 10% level. Compared to Mitaritonna et al. (2017) who use a similar setting and the Immigrant Share measurement, but find a positive coefficient. These results are however difficult to compare directly. The studies themselves differ on several levels, where the former looks at more disaggregated regions, employed natives in a region instead of natives in a firm and accounts for native outflows. However, if we look at the results for Ortega and Verdugo (2022) that does not account for native outflows, they find a positive but insignificant point estimate of 0.258% compared to 0.488% for Mitaritonna et al. (2017).

Edo (2020) provides a direct comparison between using the Immigrant Share and Migrant Inflow variable. Where they revisit an exogenous inflow of French repatriates as a result of Algerian independent in 1962. They find that their results are very similar whether or not they use the immigrant share or migrant inflow definition, where for the Immigrant Share they find a point estimate of -2.08%, and for the Migrant Inflow it is slightly less negative at -1.98%.

This therefore raises the question for the UK literature, where there are no two directly comparable studies using the same data and time period but different migration measurements. Firstly, do different measurements of immigration result in different estimates on the impact on native wages in the UK. Secondly, does the Migrant-Native ratio measurement also result in a negative bias, and does it affect how we use previous UK evidence in policy making. In the next section I explain how I will test for this directly, by estimating the impact of immigration on native wages in the UK using the same context and dataset of 2004-2019 using the APS dataset. I use a spatial approach, as this is the most common approach in the UK literature and allows me to estimate the total impact of immigration.

Table 2.2. Key Studies using the Skill-Cell or Mixed Approach to Estimate Wage Impact of Migration

Panel A: Skill-Cell Approach

Author	*Migrant Variable*	*Specification*	*Country*	*Dataset*	*Group*	*Wage Result*
Borjas (2003)	Immigrant Share	ols, weighted, decadal	USA	Census and CPS, 1960-2001	native men	-0.57(0.16)
Aydemir and Borjas (2007)	Immigrant Share	ols, weighted, decadal	Canada	Census 1971-2001	natives men	-0.51(0.2)
	Immigrant Share	ols, weighted, decadal	USA	Census, 1960-2000	natives men	-0.49(0.22)
Steinhardt (2011)	Immigrant Share	ols, weighted, decadal	Germany	IAB Subsample, 1975-2001, occ x exp	natives	-0.16(0.035)
Borjas (2014)	Immigrant Share	ols, weighted, decadal	USA	Census and ACS 1960-2011	natives men	-0.53
Bratsberg (2014)	Immigrant Share	ols, weighted, yearly	Norway	Several Admin Registers, 1998-2005	Natives	-0.278(0.175)
Llull (2014)	Immigrant Share	iv, weighted, decadal	Canada, USA	Census 1960-2000	natives men	-1.66 (0.66)
Ortega and Peri (2014)	Immigrant Share	ols, weighted, 7 yearly	France	Census, LFS, 1968-1999	native men	0.33 (0.107)
Card and Peri (2016)	Immigrant Share	ols, weighted, decadal	USA	Census and ACS, 1960-2011	native men	-0.237 (0.118)
	Migrant Inflow	ols, weighted, decadal	USA	Census and ACS, 1960-2011	native men	-0.124 (0.132)
Llull (2017)	Immigrant Share	iv, weighted, decadal	Canada, USA	Census 1960-2000	native men	-1.48(0.557)
Breunig et al. (2017)	Immigrant Share	ols, weighted, yearly	Australia	SIH, 2003-2012	natives	0.612 (0.413)
Sharpe and Bollinger (2020)	% hours worked	ols, weighted, decadal	USA	Census and ACS, 1990-2011, occ x exp	native men	-1.001(0.217)

Panel B: Mixed Approach

Author	*Migrant Variable*	*Specification*	*Country*	*Dataset*	*Group*	*Wage Result*
LaLonde and Topel (1991)	Migrant Stock	ols, weighted, decadal	USA	Census, 1970 and 1980, MSA x arrival cohort	recent arrivals	-0.09(0.03)
Card (2001)	Relative Supply	iv, weighted, cross-section	USA	Census 1990, MSA x occupation	native, men	0.1(0.03)
Borjas (2006)	Immigrant Share	ols, weighted, decadal	USA	Census, 1960-2000, MSA x duc x exp	natives	0.06(0.02)
Card and Lewis (2007)	Relative Supply	iv, weighted, decadal	USA	Census, 1980-2000, MSA x Educ	natives, men	-0.04(0.06)
Card (2009)	Relative working hours	iv, weighted, decadal	USA	Census and ACS, 1980-2006, MSA x Education	natives, men	-0.42(0.28)
Lewis (2011)	Relative Supply	iv, weighted, decadal	USA	Census, 1980-2000, MSA x Education	natives, manufacturing	-0.14(0.04)
Glitz (2012)	Relative Supply	iv, weighted, yearly	Germany	IAB Subsample, 1996-2001, region x education	natives	-0.26(0.19)
Dustmann and Glitz (2015)	% growth of skill group labour force	iv, weighted, decadal	Germany	IAB Subsample, 1985-1995, region x educ	natives, manu	-0.1(0.06)
Ozden and Wagner (2016)	Employed Migrant Stock	iv, weighted, yearly	Malaysia	LFS, 2000-2010, region x industry	natives	0.02(0.01)
Nickell and Saleheen (2017)	Migrant-Native Ratio	iv, weighted, yearly	UK	LFS, ASHE, 1992-2016, region x occ	natives	-0.082(0.091)

Notes: This table presents the results for regression estimates of various measurements of migration on log wages or earnings using the Skill-Cell Approach in Panel A, and the Mixed Approach in Panel B. Standard errors are in parentheses unless otherwise indicated, where a * indicates it was reported as a t-statistic. Due to the various differences between studies the results are not directly comparable. The main sources of data listed are as follows: ACS = American Community Survey, ASHE = Annual Survey on Hours and Earnings , CPS= Current Population Survey, DADs =Déclaration Annuelle des Données Sociales ,EAE = Enquête Annuelle d'Entreprise (an annual business survey) , FQP = Enquete Formation et Qualification Professionnelles, GHS = General Household Survey, HILDA = Household, Income and Labour Dynamics in Australia , IAB = IAB Employment Subsample, IDA = Danish Integrated Database for Labor Market Research, LFS = Labour Force Survey, WRS = Worker Registration Scheme. A regression is classified as weighted if regression weights are used on aggregated groups. MSAs= Metropolitan Statistical Areas. OLS = Ordinary Least Squares, IV = Instrumental Variables.

Table 2.3. Key Studies using the Spatial Approach to Estimate Wage Impact of Migration

Author	*Migrant Variable*	*Specification*	*Country*	*Dataset*	*Group*	*Wage Result*
Altonji and Card (1991)	Immigrant Share	iv, weighted, decadal	USA	Census, 1970-1980, 120 MSAs	native, white dropouts	-1.1(0.64)
Winter-Ebmer and Zweimüller (1996)	Immigrant Share	iv, yearly	Austria	Austrian Social Security Records, 1981-1991	native, young blue-collar	0.0025(5.1*)
Dustmann et al. (2005)	Migrant-Native Ratio	iv, weighted, yearly	UK	LFS, Census, 2003-2010	natives	0.91 (0.58)
Card (2007)	Immigrant Share	iv, weighted, cross-section	USA	Census, 1980-2000, MSA	natives	0.06(0.01)
Boustan et al. (2010)	Migrant Inflow	iv, weightd, cross-section	USA	Census, 1940, 69 MSAs	natives	0.01 (0.54)
Dustmann et al. (2013)	Migrant-Native Ratio	iv, yearly	UK	Census, LFS, 1997-2005, 17 regions	natives	0.4(0.11)
					natives, 10th pct	-0.52(0.18)
					natives, 90th pct	0.41(0.19)
					natives 75th pct	-0.005(0.223)
Lemos and Portes (2013)	Migrant Inflow	ols, weighted, yearly	UK	WRS, LFS, ASHE, 2004-2006	natives	0.246 (0.276)
Sá (2015)	Migrant Inflow	iv, yearly	UK	LFS, Census, 2003-2010	natives 25th pct	-0.187(0.164)
Basso and Peri (2015)	Migrant-Inflow	iv, weighted, yearly	USA	Census and ACS, 1960-2012, 722 CZ	natives	0.25(0.2)
Foged and Peri (2016)	Immigrant Share	iv, weighted, yearly	Denmark	IDA, 1995-2008, 97 municipalities	natives, low education	1.8 (0.64)
Breunig et al. (2017)	Immigrant Share	ols, weighted, cross-section	Australia	HILDA, Census, 2011	natives	0.264 (0.202)
Dustmann et al. (2017)	Migrant Inflow	iv, weighted, yearly	Germany	German Social Security Records, 1990-1993	Natives	-0.134(0.047)
Mitaritonna et al. (2017))	Immigrant Share	iv, weighted, yearly	France	FLFS, EAE, DADS Poste, 1995-2005	natives, manu	0.488(0.073)
Jaeger et al. (2018)	Migrant Inflow	iv, weighted, yearly	USA	Census and ACS, 1960-1980, 109 MSAs	natives	-0.193(0.117)
Edo (2020)	Immigrant Share	iv, weighted, cross-section	France	Census 1962-1968; FQP 1964-1977	Natives	-2.08(0.64)
Edo (2020)	Migrant Inflow	iv, weighted, cross-section	France	Census 1962-1968; FQP 1964-1977	Natives	-1.98(0.52)
Ortega and Verdugo (2022)	Migrant-Inflow	iv, weighted, 7-9 yearly	France	DADS, Census, 1968-2007, baseline commuter zone	native men	-0.238(0.121)
Ortega and Verdugo (2022)				DADS, Census, 1968-2007, current commuter zone	native men	0.258(0.174)

Notes: This table presents the results for regression estimates of various measurements of spatial variation in migration on log wages or earnings. Standard errors are in parentheses unless otherwise indicated, where a * indicates it was reported as a t-statistic. Due to the various differences between studies the results are not directly comparable. The main sources of data listed are as follows: ACS = American Community Survey, ASHE = Annual Survey on Hours and Earnings , CPS= Current Population Survey, DADs =Déclaration Annuelle des Données Sociales ,EAE = Enquête Annuelle d'Entreprise (an annual business survey) , FQP = Enquete Formation et Qualification Professionnelles, GHS = General Household Survey, HILDA = Household, Income and Labour Dynamics in Australia , IAB = IAB Employment Subsample, IDA = Danish Integrated Database for Labor Market Research, LFS = Labour Force Survey, WRS = Worker Registration Scheme. A regression is classified as weighted if regression weights are used on aggregated groups. MSAs= Metropolitan Statistical Areas. OLS = Ordinary Least Squares, IV = Instrumental Variables.

2.4 Methodology

In this section I set up my econometric specification, where I use a spatial approach, and discuss how I instrument for any endogeneity.

2.4.1 Empirical Strategy

The methodology builds upon the standard spatial model such as in Dustmann et al. (2013), where the authors use UK data to estimate the total effect of migration into a region on native wages across the wage distribution. However, in this paper we focus only on the average total effect on the region itself, using the following specification:

$$\triangle lnW^N_{rt} = \alpha + \beta_p \triangle m_{rt} + \triangle X_{rt} + \gamma_t + \triangle \epsilon_{rt} \tag{2.6}$$

where $\triangle lnW^N_{rt}$ is the yearly change ($lnW^N_{rt} - lnW^N_{rt-1}$) in average log native wages in region r and time t and $\triangle m_{rt}$ is the yearly change in the migration within a region r and time t. Where migration is defined in the following three ways, i) Migrant-Native Ratio $\frac{M_{rt}}{N_{rt}}$ ii) Migrant Inflow $\frac{M_{rt}-M_{rt-1}}{N_{rt-1}+M_{rt-1}}$ iii) Immigrant Share $\frac{M_{rt}}{N_{rt}+M_{rt}}$

I first difference to control for time invariant spatial factors and further control for regional characteristics X_{rt}, which includes controls for the average age for natives and migrants and education controls, defined by the age they left education, for the proportion of migrants and natives with higher (25>), high (20-24), intermediate (16-19) and low education (16 <) all within a region-year group. Lastly, I control for country-wide variation using year fixed effects, γ_t. I cluster standard errors at the region level.

One issue when allowing for spatial variation is that it is possible for natives to react to migration, for example moving to a different region. This would result in our coefficient being biased towards zero. I follow Dustmann et al. (2013) and use broad definitions of spatial regions which will reduce the likelihood of this being the case I also consider a specification using more disaggregated NUTS2 regions[4]. Furthermore, in Section 2.5.2, I discuss this issue in more detail and following the above specification, regress native outflows on three different migration variables, where I find insignificant results for both 18 Government Office Regions and 36 NUTS2 Regions.

2.4.2 Identification

A common concern when estimating the impact of migration on native wages is the endogenous allocation of migration into regions. Results would be upwardly biased if migrants move to regions experiencing high growth, that I account for by using a shift-share instrument, a standard approach in the literature. Following studies such as Bartel (1989) and Munshi (2003) the literature has utilised the findings that immigrants tend to migrate to where there is other migrants. I follow Card (2001) and construct an instrument that captures the 'supply-push' component of immigration inflows, that is those flows which are exogenous to local demand shocks and are the result of migrants moving to areas with other migrants similar to themselves [5].

To construct this instrument, I find the change in the migrant stock for the UK as a whole for each country of origin group j, $\triangle M_{jt}$[6]. I then multiply that by the share of

[4]Overall, evidence shows that internal migration to Local Authorities in England and Wales is consistent and low, increasing from 2.65 million in 2004 to 2.85 million in 2015(White, 2017). This is an overall decrease from 4.99% of the population to 4.92%. Furthermore, Chen et al. (2021) shows that internal migrants tend to move to local authorities which are nearby.

[5]Jaeger et al. (2018) show that if the distribution of the country of origin of migrants within a region remains stable overtime then using a shift-share instrument may not be sufficient. As I construct past migration shares are constructed using the 1991 Census then this is unlikely to be the case, where the Treaty of Maastricht introduced freedom of movement to the EU in 1992, where it was then subsequently expanded to ten Central and Eastern European countries in 2004 and further expanded in 2007.

[6]Where we group migrants into 10 broad regions: Republic of Ireland, Old Commonwealth, Western Europe and Cyprus , Central Europe, Turkey and Former USSR , Africa, Indian Subcontinent , Caribbean and Other America , Middle East , Other Asia, Rest of the World

migrants from each country of origin in 1991 to determine the exogenous regional distribution of migrants. Where in the absence of shocks to the local labour market, then these new migrants would distribute themselves across the UK according to where migrants from their region are already settled in the UK. For this instrument to be valid, I must assume that past regional shares are sufficiently lagged such that they are not correlated with wage growth in the region. To do so I use the 1991 Census 2% Sample of Anonymised Records to obtain the regional shares. If I let λ_{jr91} to be the share of immigrants with country of origin group j in region r in 1991, then the new number of immigrants with country of origin j that is expected to move into labour market region r is given by $\lambda_{jr91} \times \triangle M_{jt}$. This is calculated for each country of origin and then summed together. Finally, to standardise our instrument I divide it by the overall working age labour force in region r lagged three times at t-3. By lagging our labour force variable we can account for any population changes induced by higher growth rates. The advantage of using three lags over shorter lags is that it reduces the likelihood that the different population levels are correlated with higher growth rates. Where if wage growth is correlated overtime then it is possible that those groups with a higher population and wage growth in the previous one or two years also have higher population and wage growth today (Mitaritonna et al., 2017). This variable is then used in as an instrument for each of the three measurements of migration in a Two-Stage Least Squares regression.

$$SP_{jrt} = \frac{\sum_j \lambda_{jr91} \triangle M_{jt}}{L_{rt-3}} \tag{2.7}$$

Table 2.4 shows the correlation between our instrument and each measurement of migration, showing that there is a strong significant and positive correlation, although this relationship is the weakest for the Immigrant Share measurement. I then use the Olea and Pflueger (2013) weak instrument test and find an F-statistic of between 94.17-276.6 for specifications using the Migrant Inflow measurement. For the Migrant-Native ratio this is lower, between 84.85-123.3. The lowest F-stats are for the Immigrant Share measurement, which is between 14.31-75.51, but is still above the threshold of 10.

Due to the sample being split into region-time groups there is potentially an issue with measurement error. Where our sample size for migrants can be quite small in some groups, in for example regions like Northern Ireland, which can be exacerbated by first differencing our regression as discussed in Dustmann et al. (2013). However, according to the authors using an instrumental variable estimation will account for this measurement error as long as the instrumental variable's measurement error is not correlated with the measurement error of our variable of interest. As we use the 1991 Census to construct our instrument, I do not expect this to be an issue. Finally, following Dustmann et al. (2013) I do not use the APS sample weights which are calculated for the whole population, and not migrants and natives separately.

2.5 Results

In this section I discuss how spatial regression results may vary depending on the measurement of migration used. I do this for two outcomes variables, firstly for native real hourly wages and then for native outflows. For each outcome, I run the regression separately for each of the three measurements of migration, and do so firstly across 18 Government Office Regions and lastly for 36 adjusted NUTS2 regions.

2.5.1 Real Hourly Wage

The results from Table 2.5 are based on a panel dataset where for each year between 2004 and 2019 I divide the UK into 18 regions. However as I lag our instrument by three years, our sample for analysis is from 2007-2019 resulting in 234 region-time groups. As

Table 2.4. Two Stage Least Squares: First Stage Results

	18 Regions		36 Regions	
Dependent Variables	(1)	(2)	(3)	(4)
Panel A				
△ Migrant-Native Ratio	1.778***	1.750***	1.705***	1.686***
	(0.207)	(0.200)	(0.240)	(0.241)
F-stat	73.75	76.84	50.52	48.90
Effective F-Stat	119.9	123.3	89.17	84.85
Panel B				
Migrant Inflow	1.505***	1.488***	1.510***	1.491***
	(0.106)	(0.122)	(0.103)	(0.110)
F-stat	201.0	149.1	214.4	182.2
Effective F-Stat	251.6	276.6	134.4	94.17
Panel C				
△ Immigrant Share	0.525***	0.490***	0.558***	0.538***
	(0.0584)	(0.0594)	(0.0883)	(0.0837)
F-stat	80.88	68.03	39.87	41.36
Effective F-Stat	75.51	63.71	16.80	14.31
Observations	234	234	468	468
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated first-stage regression coefficients of the yearly change in our three different measurements of migration on a supply-push instrument for the years 2004-2019. Panel A shows results for the Migrant-Native Ratio measurement, Panel B the Migrant Inflow and Panel C the Immigrant Share. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects.F-stat is the Kleibergen-Paap weak instrument test and Effective F-stat is the Olea and Pflueger (2013) weak instrument test. Clustered standard errors are reported in parentheses. $***p < 0.01$ $**p < 0.05$, $*p < 0.1$

explained in the methodology section, I run the regression three times on three different measures of migration. Where Panel A is the results for the Migrant-Native Ratio in equation 2.3, Panel B is the results for the Migrant Inflow in equation 2.2 and finally the results in Panel C are for the Immigrant Share in equation 2.1. Columns 1 and 2 present the OLS results and Columns 3 and 4 our Two-Stage Least Squares (2SLS) results where the constructed Supply Push variable is the instrument. Across all three measurements of migration, I find that the results in Column 3 and 4 are more negative than in Columns 1 and 2 which is what we would expect if migrants positively selected into regions with higher wage growth. Columns 1 and 3 show the results for the specification with no controls, while Columns 2 and 4 show the results when I include controls. If I focus therefore on our preferred specification in Column 4, we can see that a 1 percentage point increase in the Migrant-Native ratio decreases native real hourly wages by -0.123%. This is compared to -0.145% for Migrant Inflow and -0.439% for the Immigrant Share. However these are only significant at a 10% level.

Our results for the Immigrant Share and Migrant Inflow definitions are as expected, where as discussed in Card and Peri (2016) we would expect the Immigrant Share to be more negative than Migrant Inflow due to the potential negative bias present. Unexpectedly, the Migrant-Native Ratio is of a similar magnitude to the Migrant Inflow.

Table 2.6 runs the same regression but using 36 adjusted NUTS2 regions from 2007-2019, resulting in there being 438 region-time groups. Focusing on Column 4, we see that for all measurements of migration the coefficients are almost double the magnitude as previously. However the same pattern persists, where the Migrant-Native Ratio is the least negative coefficient at -0.285%, followed closely by Migrant Inflow at -0.322%. The Immigrant Share continues to be the the most negative at -0.891%.

Overall, these results suggest that we should be cautious when comparing coefficient estimates from papers that use different measures of migration, however it is unlikely

Table 2.5. Impact of Three Migration Measures on Native Wages - 18 Government Office Regions

Dependent Variable	OLS		IV	
△ Log Real Hourly Wage	(1)	(2)	(3)	(4)
Panel A				
△ Migrant-Native Ratio	0.0249	0.0342	-0.120	-0.123*
	(0.0620)	(0.0663)	(0.0954)	(0.0682)
Panel B				
Migrant Inflow	-0.0546	-0.0421	-0.142	-0.145*
	(0.0854)	(0.0964)	(0.108)	(0.0803)
Panel C				
△ Immigrant Share	0.0355	0.0898	-0.407	-0.439*
	(0.158)	(0.168)	(0.308)	(0.234)
Observations	234	234	234	234
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) migrants for three different measures of migration for the years 2004-2019 over 18 government office regions. Panel A shows results for the Migrant-Native Ratio measurement, Panel B the Migrant Inflow and Panel C the Immigrant Share. Wages are winsorised at the 1st and 99th percentile. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. Clustered standard errors are reported in parentheses.
$***p < 0.01 **p < 0.05, *p < 0.1$

that previous papers in the UK that use the Migrant-Native Ratio exhibit any negative bias.

Why do we see different results?

To explain why the Migrant-Native Ratio does not appear to be biased, we should consider the discussion in Section 2.3.1. Following Card and Peri (2016), the Immigrant Share measurement should be more negative than the Migrant Inflow, which is what we see in Table 2.5 . However, we would therefore expect the Migrant-Native Ratio to also be more negative than the Migrant Inflow, when it is in fact less negative.

To explain why, lets begin by considering why the Migrant-Native Ratio does not appear negatively biased like the Immigrant Share. There are two key features that are different

Table 2.6. Impact of Three Migration Measures on Native Wages - 36 NUTS2 Regions

Dependent Variable	OLS		IV	
△ Log Real Hourly Wage	(1)	(2)	(3)	(4)
Panel A				
△ Migrant-Native Ratio	0.0198	-0.00339	-0.268	-0.285*
	(0.0967)	(0.0719)	(0.197)	(0.155)
Panel B				
Migrant Inflow	-0.165**	-0.100	-0.303	-0.322**
	(0.0741)	(0.0709)	(0.197)	(0.154)
Panel C				
△ Immigrant Share	0.0110	-0.0162	-0.820	-0.891**
	(0.165)	(0.128)	(0.545)	(0.436)
Observations	468	468	468	468
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) migrants for three different measures of migration for the years 2004-2019 over 36 adjusted NUTS2 regions. Panel A shows results for the Migrant-Native Ratio measurement, Panel B the Migrant Inflow and Panel C the Immigrant Share. Wages are winsorised at the 1st and 99th percentile. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. Clustered standard errors are reported in parentheses.
$***p < 0.01 **p < 0.05, *p < 0.1$

in the Taylor expansion.

Firstly, there is a difference in weighting the migrant inflow component of each equation. Where in equation 2.4, this component is weighted by $1 - m^{ML}_{rt-1}$. Whereas, in equation 2.5, we can see that this term is given a weight of 1. This means, if we assume that N_{rt-1} and M_{rt-1} remain constant, then the larger $\triangle M_{rt}$ then the larger the change in Migrant-Native Ratio, $\triangle m^{MN}_{rt}$, will be compared to the change in the Immigrant Share, $\triangle m^{ML}_{rt}$. Furthermore, compared to the Migrant Inflow, $\triangle m^{MF}_{rt}$, this term will in fact be larger for the change in $\triangle m^{MN}_{rt}$ due to the smaller denominator N_{rt-1} as opposed to $N_{rt-1} + M_{rt-1}$.

Secondly, lets consider the native inflow component of each measurement, which is likely to be endogenous, resulting in a negative bias to our estimates. The native inflow term will receive a higher weight when using $\triangle m^{MN}_{rt}$ compared to $\triangle m^{ML}_{rt}$. This is because given the same N_{rt-1} and M_{rt-1}, then the lagged Migrant-Native Ratio m^{MN}_{rt-1} is larger than the lagged Immigrant Share m^{ML}_{rt-1}. Furthermore, for a constant number of N_{rt-1}, as M_{rt-1} increases, m^{MN}_{rt-1} would increase by more than m^{ML}_{rt-1}, resulting in a larger negative bias from endogenous native inflows in regions where m^{MN}_{rt-1} is larger than m^{ML}_{rt-1}. As such, we may expect the Migrant-Native Ratio to have a smaller negative bias than the Immigrant Share, so long as the difference between the first component is large enough.

Considering these two channels, there is also a scenario where the Migrant-Native Ratio is not negatively biased and may even be positively biased compared to the migrant inflow. The condition that must be met is that the difference between the change in the number of migrants as a proportion of the lagged number of natives for the change in the Migrant-Native Ratio and the Migrant Inflow is more than or equal to the negative bias from endogenous native inflows in the change in the Migrant-Native Ratio measurement.

Condition 1:

$$\frac{\triangle M_{rt}}{N_{rt-1}} - \frac{\triangle M_{rt}}{N_{rt-1} + M_{rt-1}} \geq m^{MN}_{rt-1} \frac{\triangle N_{rt}}{N_{rt-1}}$$

In this scenario, the larger the endogenous change in the number of natives as a proportion of the native population, then the more difficult it will be to meet this condition. However, assuming this is fixed, then this condition is easier to meet depending on two related factors. Firstly, the larger the lagged number of migrants, M_{rt-1}, the larger the difference between the change in the number of migrants as a proportion of the native labour force $\frac{\triangle M_{rt}}{N_{rt-1}}$, and the Migrant Inflow. Secondly, however, if M_{rt-1} is larger, then so would the lagged Migrant-Native Ratio, m^{MN}_{rt-1}, which would increase the weight given to endogenous native inflows. Which term dominates is likely to also depend on how large the lagged local native labour force is, and how large the change in natives are relative to the change in migrants for a given lagged number of natives, assuming that there are no region-time cells where the number of migrants is larger than the number of natives. This could explain why in Table 2.5 and 2.6, the Migrant-Native Ratio is less negative than the Migrant Inflow.

2.5.2 Native Outflows

When using spatial variation to estimate the impact of migration on native wages, it is possible that in response to immigration into a region, natives will respond by leaving the region. This would spread the local supply shock across regions and reduce the potential impact from immigration (Borjas, 2006, Sá, 2015). The evidence of this occurring in the UK is mixed. However, similarly to our wage regressions, Card and Peri (2016) show that estimates that using the Immigrant Share to estimate the impact of migration on native outflows will be negatively biased, which would also be the case for native outflows.

Furthermore, estimates that use the Immigrant Share or Migrant-Native ratio could produce a stronger negative bias than in wage estimations where the second term in equations 2.4 and 2.5 is mechanically negatively correlated to native outflows. In the UK the evidence for this relationship is mixed. Dustmann et al. (2013) who use LFS data for 17 UK regions, excluding Northern Ireland, from 1997-2007 find no evidence of this occurring. However, Sá (2015) who use LFS data for 170 local authorities from 2003 to 2010 in England and Wales, and use the Migrant Inflow find a significant negative effect of -0.868. This difference is likely explained by the level of regional aggregation as opposed to the measure of migration. To investigate this I run a regression similar to that of our main specification laid out in equation 2.6, except for our outcome variable we use the net native outflow normalised by the lagged labour force: $\frac{\triangle N_{rt}}{L_{rt-3}}$. As it is reasonable to assume that both natives and migrants would want to move into the same areas with high growth, I follow Sá (2015) in instrumenting our regression using the supply push instrument.

Table 2.7. Impact of Three Migration Measures on Native Outflows - 18 Government Office Regions

Dependent Variable	OLS		IV	
△ Native Outflows	(1)	(2)	(3)	(4)
Panel A				
△ Migrant-Native Ratio	-0.531	-0.360	-0.0303	0.263
	(0.411)	(0.375)	(0.544)	(0.604)
Panel B				
Migrant Inflow	1.620	1.657	-0.0338	0.285
	(0.986)	(1.053)	(0.607)	(0.638)
Panel C				
△ Immigrant Share	-1.279	-0.975	-0.0964	0.876
	(0.905)	(0.889)	(1.731)	(2.001)
Observations	216	216	216	216
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) native outflows on the yearly change in the working age(16-64) migrants for three different measures of migration for the years 2004-2019 over 18 government office regions. Panel A shows results for the Migrant-Native Ratio measurement, Panel B the Migrant Inflow and Panel C the Immigrant Share. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. Clustered standard errors are reported in parentheses. $***p < 0.01$ $**p < 0.05$, $*p < 0.1$

Tables 2.7 and 2.8 follow the same structure the previous regressions. Where Panel A is the results for the Migrant-Native Ratio in equation 2.3, Panel B is the results for the Migrant Inflow in equation 2.2 and finally the results in Panel C are for the Immigrant Share in equation 2.1. Columns 1 and 2 present the OLS results and Columns 3 and 4 our Two-Stage Least Squares (2SLS) results where the constructed Supply Push variable is the instrument. The OLS results in Columns 1 and 2 of Tables 2.7 and 2.8 follow a similar pattern to that found in Card and Peri (2016). Where the Immigrant Share is mechanically negatively related with native outflows, then it has a strong negative coefficient. This is also the case for the Migrant-Native Ratio which we showed also has a mechanically negative relationship with native outflows. In contrast to this, the Migrant

Table 2.8. Impact of Three Migration Measures on Native Outflows - 36 NUTS2 Regions

Dependent Variable	OLS		IV	
$\triangle$ Native Outflows	(1)	(2)	(3)	(4)
Panel A				
$\triangle$ Migrant-Native Ratio	-0.624	-0.722	-0.426	-0.0585
	(0.460)	(0.372)	(0.953)	(0.791)
Panel B				
Migrant Inflow	2.088*	1.643*	-0.447	-0.0607
	(0.994)	(0.703)	(1.030)	(0.825)
Panel C				
$\triangle$ Immigrant Share	-0.968	-1.157	-1.262	-0.176
	(0.811)	(0.666)	(2.820)	(2.390)
Observations	432	432	432	432
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) native outflows on the yearly change in the working age(16-64) migrants for three different measures of migration for the years 2004-2019 over 36 adjusted NUTS2 regions. Panel A shows results for the Migrant-Native Ratio measurement, Panel B the Migrant Inflow and Panel C the Immigrant Share. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. Clustered standard errors are reported in parentheses. $***p < 0.01$ $**p < 0.05$, $*p < 0.1$

Inflow measurement has a large positive correlation, this is likely driven by natives and migrants moving to regions with high wage growth which is not fully captured by the other two migration measurements due to the negative bias present.

In Table 2.7, once I instrument for migration correlated with wage shocks in Columns 3 and 4, we can see in Column 3 the coefficient is essentially zero for all measurements of migration, but once I add controls this becomes positive for all specifications. However, the IV results are extremely imprecise and should be interpreted with caution. Table 2.8 shows a different pattern. Where in Column 3 the results for Migrant-Native Ratio and the Immigrant Share remain negative, and Migrant Inflow becomes negative also. Whereas when we add controls in Column 4, the Migrant Native ratio and Migrant Inflow are essentially zero, while the Immigrant Share remains negative but with a greatly reduced magnitude. The results must continue to be interpreted with caution as they are extremely imprecise.

As expected, we can see that in the specification with 36 regions, the coefficients are more negative for all specifications. If I focus on Column 4, this suggests that when 100 migrants enter one of the 18 government office regions this could result in 26 natives moving in also. Whereas, when 100 migrants move into one of the 36 NUTS2 regions, this would result in a decrease of about 5 natives. One potential reason for this is that when migrants move into a NUTS2 region, this may attract natives from further away in the UK who want to reside in areas with more migrants. However, those natives living in the NUTS2 region, may decide to move to a NUTS2 region that is within the same government office region and as a result in our 36 region specification these two mechanisms cancel each other out.

2.6 Conclusion

How migration is measured can bias the results of regressions estimating the impact of native wages. In this paper, I used UK Annual Population Survey data to explore three common measures of migration. Following Card and Peri (2016), I showed that both the Immigrant Share and the Migrant-Native Ratio should produce negatively biased estimates. I further showed that since Card and Peri (2016), the literature has paid more attention to how they measure migration, however due to the variation in settings and approaches it is not possible to determine to what extent the difference in results can be explained by the measure of migration, where only Edo (2020) compares the Immigrant Share and Migrant-Native Ratio directly for an exogenous inflow of French repatriates in 1962, finding little difference in point estimates. Similar to Card and Peri (2016), I find that the Immigrant Share measurement of migration obtains a more negative estimate than the unbiased Migrant Inflow variable. I also find that the Migrant-Native Ratio measurement unexpectedly obtained a similar estimate to the Migrant Inflow. I show that this will likely hold under certain underlying conditions in the local labour markets. Where for a given level of endogenous native inflows, it is dependent on the lagged number of migrants, and the lagged migrant-native ratio. Finally, I show that for all measures of migration, there is no significant impact on native outflows, which could positively bias our wage regression.

These results show the importance of being careful when measuring the impact of migration on native wages, and policymakers using past studies on migration to make decisions should consider how the measurement of migration may effect the results.

Chapter 3

Cross-Occupational Effects of Immigration on Native Wages in the UK[1]

[1] This chapter is based upon and extends research that appears in an earlier discussion paper co-authored with my PhD supervisors, Dr Marco Alfano and Prof Graeme Roy, which is available in Alfano, Mckenzie, and Roy (2020).

3.1 Introduction

The impact of immigration on the wages of natives remains a topic of intense debate, both in economic policy circles and the wider public discourse. The magnitude and even direction of this effect appears to vary based on setting and approach taken (Dustmann et al., 2016). Studies typically investigate whether natives and migrants either compete with or complement each other in similar jobs or skill groups, often referred to as cells. There is a rich body of evidence looking at whether or not migrants either compete with or complement natives in the same part of the wage distribution—i.e. within the same cell. However, whether or not these same migrants yield benefits or costs to native workers just above or below them in the wage distribution—i.e. in an adjacent cell—has remained relatively unexplored. These cross-effects could be the result of migrants producing positive productivity spillovers from peer effects or allowing for natives to specialise. This is important as when determining migration policy, policymakers should consider it's effect on natives not only in the same part of the wage distribution.

In this paper, we estimate the effect of immigration into an occupation on wages of natives working in higher paid occupations. Such *cross-occupational* effects of immigration may arise by migrants increasing the productivity of workers (Peri et al., 2015, Ottaviano et al., 2018 for instance) or by migrant inflows allowing natives to specialize in more complex, better remunerated tasks (Peri and Sparber, 2009; for example). Such effects may be even more likely in countries such as the UK, where migrants have been found to downgrade upon arrival thus leading to an inflow of over-qualified workers (Dustmann et al., 2013). Whilst we do not detect any meaningful effect of immigration within the same occupation-region group, we find that immigration into one occupation increases wages of natives working in the occupation ranked above by around 0.332 percent. Our findings are consistent with migrants increasing productivity and allowing natives to specialise, however we only find weak evidence of these channels being present.

To estimate the effect of immigration into an occupation on wages of natives working in higher paid occupations we use Office for National Statistics (ONS) data and divide workers in each of the 13 U.K. regions into 9 occupational categories based on the Standard Occupational Classification (SOC). To identify adjacent occupations, we first rank all 9 occupations according to the ordering provided by the SOC 2010 ordering with Managers, Directors and Senior Officials at the top and Elementary Occupations at the bottom. For each occupation o, we define the occupation *below* ($o - 1$) as the occupation with mean hourly earnings are one rank lower than o. Similarly, the occupation *above* ($o + 1$) is the occupation with mean hourly earnings one rank higher than o. Using these definitions, we regress yearly changes in native wages in occupation o on yearly changes the migrant-native ratio in occupations o, $o - 1$, $o + 1$. As such, this paper builds upon Dustmann et al. (2013) in trying to identify the underlying cross-effects of migration within these regions. Following standard practice in the literature, we instrument migration flows using the supply-push instrument first detailed in Card (2001).

We first describe the occupational distribution of migrants and find that migrants tend to cluster at the top and the bottom of the wage distribution. Moreover, in later years, migrants are increasingly found working at low paying occupations. When we use the methodology adopted by Dustmann et al. (2013) we find positive and significant effects of a very similar magnitude to the authors.

To estimate cross-occupational effects in the same setting, we construct an occupation-region-year panel. We find that wages of natives working in occupation o are increased by immigration into the occupation *below*. Our point estimates suggest that a 1 percent increase in the change in the migrant-native ratio in occupation $o - 1$ results in a 0.332 percent increase in native wages in occupation o. By contrast, we find no effects of immigration into the same occupation (o) or into the occupation above ($o + 1$). The results for migration into the same occupation and below occupation are robust to changes in occupation orders, where we instead order occupations by the average real hourly wage. However, when we order according to average real hourly wage, we find a significant but negative effect from migration into the above occupation. Nevertheless, after calculating the average yearly change in migration in below and above occupations and multiplying it by their corresponding coefficients, we find for both specifications that the average yearly effect of migration from below and above is positive overall. Moreover, we find that the positive wage effect from migrants working in occupations *below* natives is concentrated in occupations located at the lower end of the wage distribution. However, likely due to a smaller sample size, our results are insignificant.

We consider productivity, peer effects and specialisation as possible channels of impact. Peer effects may impact productivity and therefore native wages as a result of social pressure to work harder and/or through knowledge spillovers (Cornelissen et al., 2017). It is reasonable to assume that migrants would not only interact with those within their own occupation and so it is possible that these spillovers could occur across occupations. To provide evidence on this mechanism we first show that higher migrant inflows into an occupation are associated with higher levels of education in that same occupation. This is partly a consequence of the fact that migrants' educational attainments exceed that of natives and in part due to migrants downgrading upon arrival to the UK (Dustmann et al., 2013). In a second step, we regress wages of natives within occupation o, with the average educational attainments of employees working in the occupation below, $o-1$, and find a weak but positive correlation. Taken together, these two regressions provide weak evidence that migrants may lead to *cross-occupational* wage impacts due to their exceptionally high levels of education and warrants further exploration in future studies.

An alternative pathway of impact through which migrants increase the wage of those natives working in occupations above their own is by allowing natives to specialise in better paid tasks. Peri and Sparber (2009) show when migrants have a comparative advantage in 'lower' skilled, manual occupations then natives are pushed to specialise in 'higher' skilled occupations with complex communicative, interactive and better remunerated tasks. As a result, we expect that when migrants enter into an adjacent below occupation, natives can specialise in the occupation above for which they may have a comparative advantage in, and therefore receive higher wages. Due to data limitations, we investigate this channel of impact by focusing on in-job training received by natives which acts as a proxy for specialisation. Training is a likely pre-requisite for specialising in more complex tasks and accordingly we find that migration flows into occupation o induce natives in the same occupation, o, to take up in-job training. This suggests that task specialising may be concentrated within broad occupation groups. These results tally with findings by Campo et al. (2018), who find that immigration is associated with higher native training, albeit only at a regional level. Although we find no significant cross-occupational effects, this warrants further exploration using a task based approach as used in previous studies.

By allowing immigration into one section of the labour force to affect natives in different occupations, this study provides evidence on a novel impact of migrants on native wages. As such, our results complement the large literature on the effect of immigration on native labour market outcomes. Dustmann et al. (2016) categorised this literature into three key methodological approaches which would drive the disparity in results throughout the literature. Firstly, the national skill cell approach which uses variation across skill-experience groups and identifies the relative wage effect of immigration by experience within a skill group and tends to find significant negative results (Borjas, 2003, 2014, Aydemir and Borjas, 2007, Ortega and Peri, 2014, Card and Peri, 2016, Llull, 2017, Sharpe and Bollinger, 2020). Secondly, the spatial approach which uses regional variation in migration and measures the absolute effect on native wages on a particular skill group and this finds a variety of results results which can be negative, insignificant or positive depending on the context (Card, 1990, Altonji and Card, 1991, Lemos and Portes, 2013, Dustmann et al., 2013, Sá, 2015, Dustmann et al., 2017, Foged and Peri, 2016, Mitaritonna et al., 2017, Jaeger et al., 2018, Peri and Yasenov, 2019, Edo, 2020, Ortega and Verdugo, 2022). Recent literature using this approach has found that migrant outflows have no positive effects on natives labour outcomes Clemens et al. (2018a), Lee et al. (2019). Thirdly, the mixed approach which uses both spatial and skill variation and measures the relative impact of migration on native wages across skill groups and finds overall either a small negative effect or no effect (LaLonde and Topel, 1991, Card, 2001, 2009, Borjas, 2006, Lewis, 2011, Glitz, 2012, Dustmann and Glitz, 2015, Nickell and Saleheen, 2015, 2017). This paper utilises a variation on the mixed-approach to investigate cross-occupational effects of migration.

Our paper also contributes to the literature on the mechanisms through which migrants affect native outcomes. By considering cross-occupational effects, we are the first to explore whether migrants that do not work in the same occupation as natives can result in the previously highlighted mechanisms arising. Previous studies have highlighted many reasons that migration can increase productivity including diversity (Ottaviano and Peri, 2006, Kerr and Lincoln, 2010, Ortega and Peri, 2014, Peri et al., 2015, Kemeny and Cooke, 2018), cost-reduction (Ottaviano et al., 2013) and bilateral trade(Gould, 1994, Rolfe et al., 2013, Ottaviano et al., 2018). Recent studies find migration increases labour productivity within firms in the UK (Ottaviano et al., 2018) and within UK regions (Campo et al., 2018). In this paper we focus on two other channels, peer effects and native specialisation. We apply previously highlighted rationales for peer effects affecting wages through productivity spillovers (Cornelissen et al., 2017), to migrant peers across occupations, which is in line with the wider literature on workplace productivity (Mas and Moretti, 2009, Falk and Ichino, 2006, Waldinger, 2012, Azoulay et al., 2010, Jackson and Bruegmann, 2009). Next, we discuss how migration may result in natives specialising in more complex and interactive tasks, which is widely explored in the literature (Peri, 2012, D'Amuri and Peri, 2014, Bisello, 2014, Foged and Peri, 2016). However, we find that this association is stronger within broad occupation groups as opposed to across them.

The remainder of the paper is as follows: Section 3.2 describes the data sources. In section 3.3 we define how we order occupations, our empirical specification and identification strategy. We then show how we will investigate the mechanisms behind spillovers. In Section 3.4 we discuss our estimation results and discuss further robustness checks in Section 3.5. Finally we discuss potential pathways of impact in section 3.6. Section 3.7 concludes.

3.2 Data, Measurements and Descrpitive Statistics

To estimate the effect of changes in migrant stock within a particular occupation on native wages in other, related occupations, we use data from the UK Annual Population Survey (APS) from 2004-2017. Using the Standard Occupational Code system (SOC) provided by the APS we divide employees into nine occupations and rank these nine occupations by the mean real hourly earnings of their employees. For each occupation o we then estimate whether changes in the migrant stock in occupations *below* and *above* occupation o have an effect on natives working in occupation o.

3.2.1 Data

We use the Annual Population Survey (APS), which provides detailed data on labour outcomes and migration for a large, representative sample for the UK with boosted samples for smaller regions. The APS consists of repeated cross sections and contains year data for the years 2004 to 2017. The APS is a survey of private households in the UK conducted by the Office of National Statistics (ONS) in Great Britain and by the Northern Ireland Statistics and Research Agency(NISRA) in Northern Ireland.[2] The sample size of the APS is made up of around 320,000 households in each survey, which is the widest ranged household survey in the UK. It allows the generation of statistics for smaller UK regions, as it utilises sample boosts from the Local Labour Force Survey and APS boost in 2004 and 2005. These local boosts allow us to break down the data to regional levels while maintaining a good sample size and accuracy. The APS contains data on employment, unemployment, income as well as information on age, education and occupation. Details about the sampling employed by the APS are reported in appendix B.1.

[2] Office for National Statistics, Social Survey Division (2021). *Annual Population Survey, 2004-2021: Secure Access* [data collection]. 21st Edition. UK Data Service. SN: 6721, DOI:10.5255/UKDA-SN-6721-7

3.2.2 Measurements

We define *migrants* as those individuals interviewed by the APS that were not born in the UK. The APS records gross weekly wages and total hours worked per week. Using these two pieces of information we calculate gross hourly wages, which we deflate by the 2015 CPI.

We construct an occupation-region-year panel for the years 2004 to 2017 by aggregating wages for those who are of working age, between the ages of 16 to 64. We divide the UK into 13 regions, 10 regions in England (Northeast, Northwest, Merseyside, Yorkshire & Humberside, East Midlands, West Midlands, Eastern, London, Southeast and Southwest) as well as Wales, Scotland and Northern Ireland. We allocate workers into 9 occupations by using the 1-digit SOC 2010 definition as follows: i) managers, directors and senior officials; ii) professionals; iii) associate professional and technical; iv) administrative and secretarial; v) skilled trades; vi) caring, leisure and other services; vii) sales and customer service; viii) process, plant and machinery; and ix) elementary occupations.

In order to sort all observations into the 1-digit SOC 2010 occupation we must convert observations in years 2004-2010 from their SOC 2000 definition to their SOC 2010 definition. We do this using probabilistic matching, utilising the dual coded APS 2011 dataset where observations are assigned both a SOC 2000 and a SOC 2010 code. The changes in codes between SOC 2000 and SOC 2010 were made at a 4-digit code level, however we only have codes up to the more aggregated 3-digit level and as such we must predict which observations would be sorted into a different occupation. Following Goos and Manning (2007) we use an unconditional matching approach. Where we use the dual-coded 2011 dataset to find the proportion of observations in each 3-digit SOC 2000 occupations that were in each 1-digit SOC 2010 occupation. Although conditional approaches have been used by Salvatori (2018), they find this had no meaningful effect on their results. The observations from 2004-2010 are then randomly sorted into a 1-digit SOC 2010 according to the proportions found in the 2011 dataset, such that it replicates the occupation distribution found in 2011. Appendix B.2 provides further information, showing that key characteristics for 1-digit SOC 2000 and SOC 2010 occupation groups in the 2004-2010 period are very similar. Furthermore, we show that there is also little difference between SOC 2000 groups immediately before and after the change in definition in 2010 and 2011, and this is also the case for SOC 2010 groups. Following standard practice in the literature, we instrument country-specific migrant shares utilising the past regional share of migrants from each country of origin in the 2% Sample of Anonymised Records for the 1991 CENSUS. [3] [4] More information on how this was used can be found in Section 3.3.

3.2.3 Summary Statistics

Table 3.1 reports selected characteristics for natives and migrants working for the years 2004 and 2017. whilst in 2004 real hourly wages of migrants exceeded those for natives, the opposite is true for the last year of our analysis, 2017. Across both time periods, migrants in work are slightly younger compared to natives, it is the case that the average age of natives increases by 1.3 years, while migrants has remains almost the same. The proportion of women working is also slightly higher for migrants. In terms of education, working migrants are –on average– better educated compared to working natives. Whilst average educational attainment improves from 2004 to 2017, the gap between natives and migrants remains relatively constant.

[3] Office for National Statistics, Census Division, University of Manchester, Cathie Marsh Centre for Census and Survey Research (2013). *Census 1991: Individual Sample of Anonymised Records for Great Britain (SARs)*[data collection]. UK Data Service 7210,DOI: 10.5255/UKDA-SN-7210-1

[4] Office for National Statistics, Census Division, University of Manchester, Cathie Marsh Centre for Census and Survey Research (2013). *Census 1991: Individual Sample of Anonymised Records for Northern Ireland (SARs)*[data collection]. UK Data Service 7212,DOI: 10.5255/UKDA-SN-7212-1

Table 3.1. Summary statistics for working age natives and migrants in 2004 and 2017

	Natives		Migrants	
	2004 Mean	2017 Mean	2004 Mean	2017 Mean
Real Hourly Wage	13.12	13.96	14.38	13.70
Age	40.91	42.22	39.57	39.76
Female	0.52	0.52	0.54	0.55
Education				
Higher(>25)	0.01	0.02	0.07	0.10
High (20-24)	0.15	0.23	0.29	0.37
Intermediate (16-19)	0.59	0.60	0.45	0.40
Low (<16)	0.19	0.08	0.11	0.06
None/Still in Education	0.06	0.07	0.08	0.08

Notes: Entries are for working age(16-64) natives and immigrants for the average real hourly wage, average age, share of female and the share in each education group in 2004 and 2016. Higher education: left full-time education after age 25, High education: left full-time education between age 20-24, Low education: left full-time education between age 16-19, None/Still in education: left education at age 15 or below, or is still in education. N is the number of observations for each statistic. Source: APS 2004, 2017

In figure 3.1 we show the proportion of those employed in an occupation who are migrants in 2004 and 2017. When compared to the UK average (shown as a red horizontal line), migrants tend to work at both the high and the low end of the occupational distribution. From 2004 to 2017, however, we see a compositional shift towards occupations on the low end of the occupation distribution. This is mainly driven by increases in migrants working in Proccess, Plant and Machinery Occupations and Elementary Occupations. Whereas the proportion of migrants among workers in Process, Plant and Machinery Occupations in 2004 was 7% (below the UK average), this occupation reports the second highest migrant share in 2017, 22%. Elementary Occupations follow a similar pattern, with increases from 8% to 23%. By contrast, the proportion of migrants in the two highest earning occupations (managerial and professionals) decreases in 2017 relative to the average for the whole workforce.

This compositional shift into occupations at the lower end of the occupational distribution could perhaps explain why average wages for migrants have fallen relative to natives despite the large increase in education. These results fit with Salvatori (2018), who finds that between 1979 and 2012 in the UK relative to natives, migrants increased the employment share in bottom paid occupations.

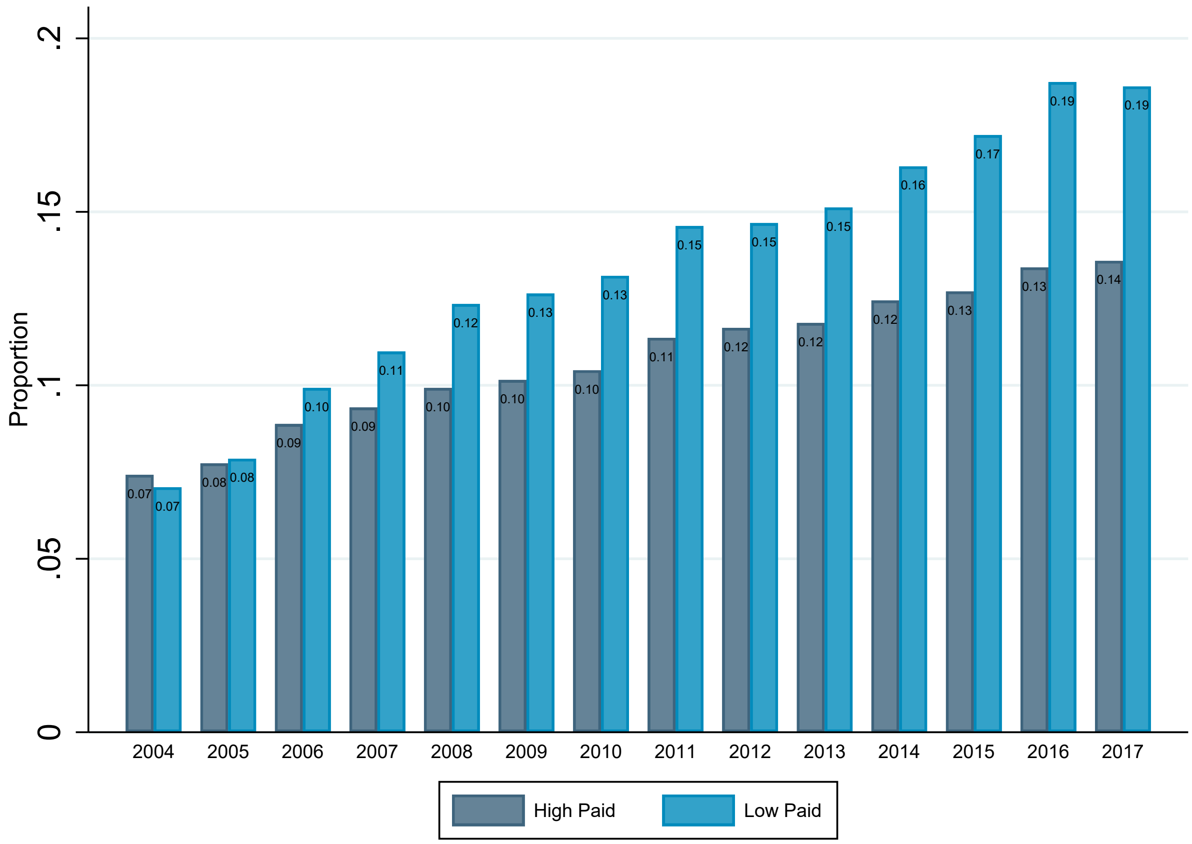
In Figure 3.2 we group occupations into Low Paid (Caring, Leisure and Other Services; Sales and Custmer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations) and High Paid (Managers, Directors and Senior Officials; Professionals; Associate Professional and Technical; Administrative and Secretarial; Skilled Trades) according to whether they were below or above the median average wage across occupations. This figure shows that the proportion of high and low paid occupation's employed workforce that are migrants has steadily increased every year, but has increased more quickly for low paid occupations relative to high paid. Where in 2004, the proportion for high and low paid was around 7%, however by 2011 the was a 4 percentage point gap, at 11% for high paid, and 15% for low paid, increasing further by 2017 to a 5 percentage point gap, at 14% for high paid and 19% for low paid.

Figure 3.3 shows the proportion of highly educated workers (defined as individuals who left full time education from age 20 and above) for each of the 9 occupations for migrants and natives in 2004 and 2017. When compared to the UK average (shown as a red horizontal line), the figure shows that migrants are better educated compared to natives across all occupations. Over time these differences increased, especially for workers employed in lower paid occupations. In fact, migrants working in elementary occupations show higher educational attainments than the UK average for all occupations. In elementary occupations in 2017, 33% of migrants are highly educated compared to just 7 % of natives. Compared to 2004, this corresponds to an increase of 13 percentage points for migrants compare to 4 percentage points for natives.

Figure 3.1. Proportion of Occupation who are migrants in 2004 and 2017

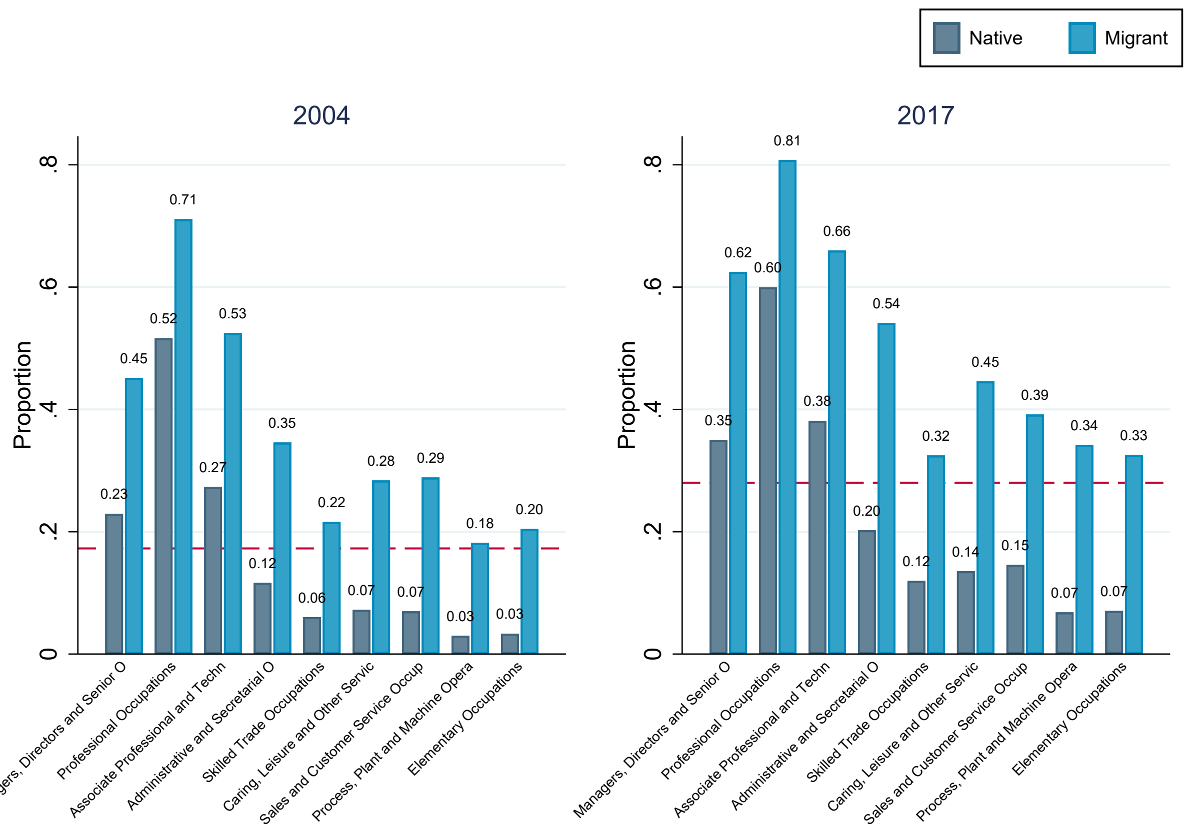
Notes: This figure shows the proportion of those employed in an occupation who are migrants in each occupation in 2004 and 2017. Where occupations are defined by the 9 1-digit SOC 2010 definition. Observations from 2004 were randomly sorted into SOC 2010 occupations based on the proportion of observations from 3 digit SOC 2000 occupations in 1 digit SOC 2010 occupations in the dual coded 2011 APS . The red line indicated the average proportion of the employed working age sample who are migrants in each respective year. Source: APS 2004-2017.

Figure 3.2. Proportion of high paid and low paid occupations who are migrants from 2004-2017



Notes: This figure shows the proportion of working age migrants in high and low paid occupations from 2004-2017. Where occupations are defined by the 9 1-digit SOC 2010 definition. Low paid occupations are defined as those below the median wage across all 1 digit SOC 2010 occupations which are: Caring, Leisure and Other Services; Sales and Customer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations, where the remaining occupations are defined as high paid. Where observations from 2004-2010 were randomly sorted into SOC 2010 occupations based on the proportion of observations from 3 digit SOC 2000 occupations in 1 digit SOC 2010 occupations in the dual coded 2011 APS. Source: APS 2004-2017.

Figure 3.3. Proportion of high and higher educated migrants and natives in each occupation in 2004 and 2017



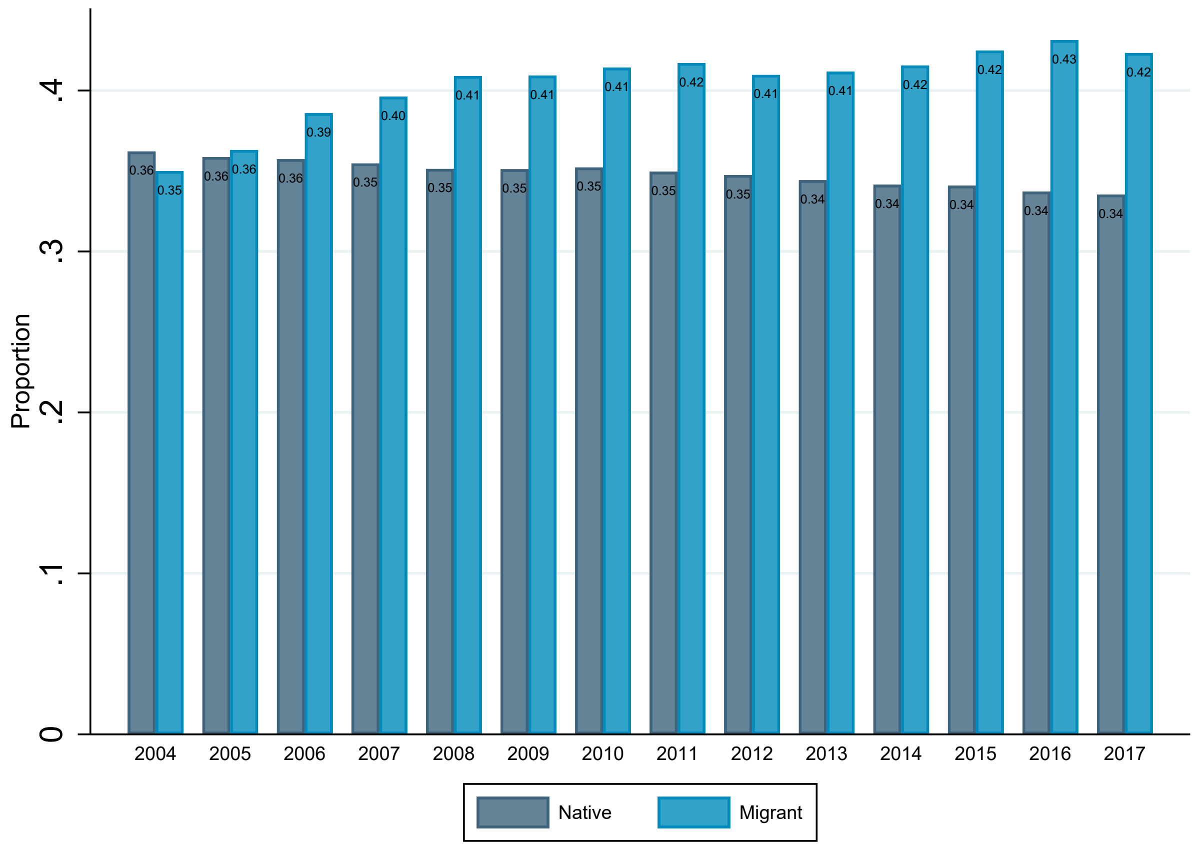
Notes: This figure shows the proportion of migrants and natives of working age (16-64) with high or higher education in each of the 9 1-digit SOC 2010 occupations. . Where observations from 2004 were randomly sorted into SOC 2010 occupations based on the proportion of observations from 3 digit SOC 2000 occupations in 1 digit SOC 2010 occupations in the dual coded 2011 APS .The red line indicates the average proportion of high and higher education who are of working age in each respective year. Source: APS 2004-2017.

Figure 3.4 shows the proportion of working age natives and migrants in low paid occupations from 2004-2017. The figure shows that in 2004, 36% of natives were in low paid occupations compared to 35% of migrants, but from 2005 onwards, there was a higher proportion of migrants in low paid occupations compared to natives. From 2005-2011, the gap widened, where the proportion of natives in low paid occupations fell slightly to 35% , while the proportion of migrants in low paid occupations increased to 42%. From 2012 this trend somewhat stagnated, peaking again in where the proportion of natives fell to 34% and the proportion of migrants increased to 43%.

In Figure 3.5 we further split migrants into older and recent migrants. Recent migrants are defined as those who arrived within 3 years of the survey date, and older migrants arrived over 3 years before the survey date. This shows that from 2004-2011 there was a large increase of the share of recent migrants entering low paid occupations, increasing by 8 percentage points from 49% to 57%. However, after 2012 this proportion dropped to 51% and by 2017 it was 53%. Whereas for natives it stayed relatively stable, at 36% in 2004 and 34% in 2017. In 2004, we can see that older migrants tended to have a smaller share in low paid occupations compared to natives at 33% and 36% respectively. This could be for two reasons, firstly that migrants tend to downgrade when they first enter the UK. Secondly, due to the expansion of freedom of movement in the EU, this resulted in a higher inflow of workers into low paid jobs. By 2017, 41% of older migrants are in low paid jobs, whilst natives fell to 34%, this suggests that overtime, either migrants downgrade for longer, or there has been a larger inflow of migrants who enter low paid jobs and are not downgrading.

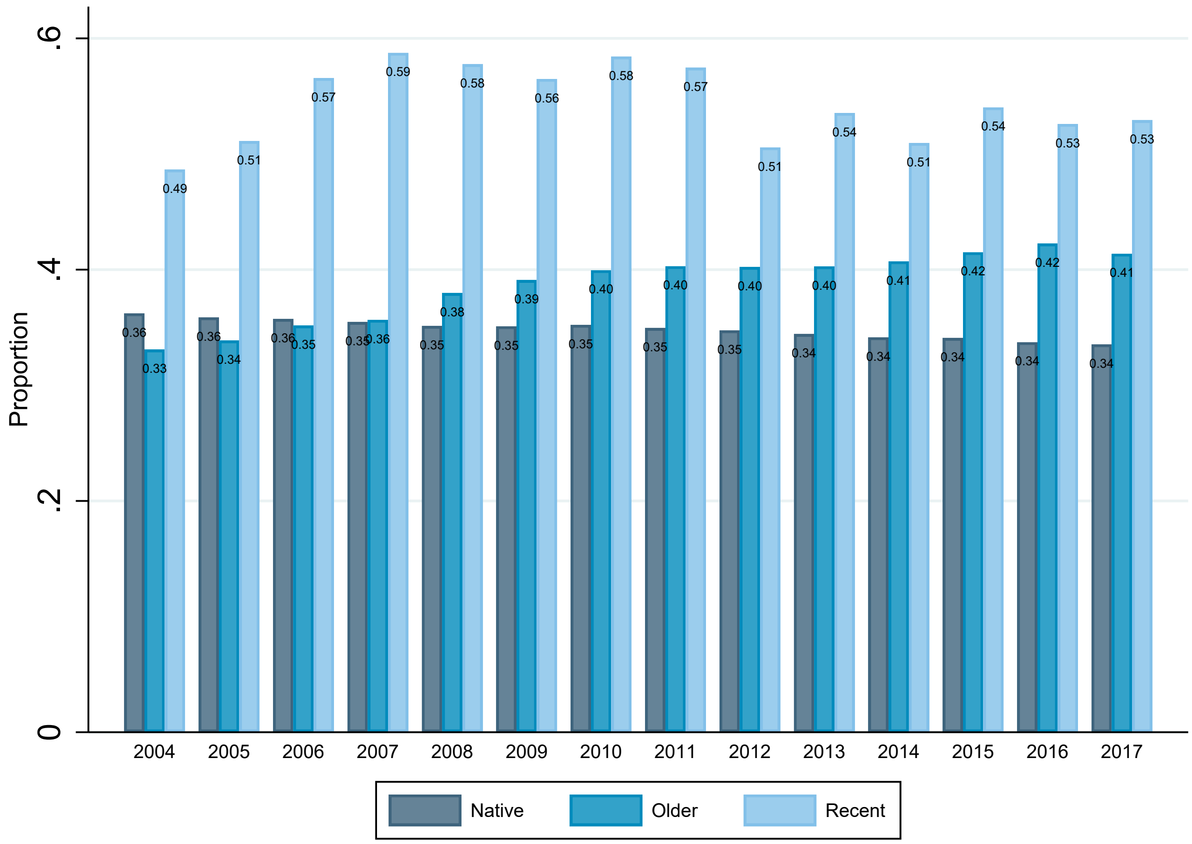
Overall, these graphs show that overtime more migrants who enter the UK work in low paid occupations, despite rising education levels over the same period. This downgrading is more likely to be the case for recent migrants, but in recent years a higher proportion of older migrants tend to remain in low paid occupations. This pattern is consistent with results presented by Dustmann et al. (2013) where the authors show that many migrants concentrate initially at the low end of the wage distribution.

Figure 3.4. Proportion of natives and migrants in low paid occupations from 2004-2017



Notes: This figure shows the proportion of working age(16-64) natives and migrants in low paid occupations from 2004-2017. Where low paid occupations are defined as those below the median wage across all 1 digit SOC 2010 occupations which are: Caring, Leisure and Other Services; Sales and Customer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations. Where observations from 2004-2010 were randomly sorted into SOC 2010 occupations based on the proportion of observations from 3 digit SOC 2000 occupations in 1 digit SOC 2010 occupations in the dual coded 2011 APS. Source: APS 2004-2017.

Figure 3.5. Proportion of natives, older migrants and recent migrants in low paid occupations from 2004-2017



This figure shows the proportion of working age(16-64) natives, older migrants and recent migrants in low paid occupations from 2004-2017. Recent migrants are defined as those who arrived within 3 years of year of the survey, and older migrants over 3 years. Low paid occupations are defined as those below the median wage across all 1 digit SOC 2010 occupations which are: Caring, Leisure and Other Services; Sales and Customer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations. Where observations from 2004-2010 were randomly sorted into SOC 2010 occupations based on the proportion of observations from 3 digit SOC 2000 occupations in 1 digit SOC 2010 occupations in the dual coded 2011 APS.

3.3 Methodology

To lay out our empirical strategy, we first discuss how we rank occupations and then set up our econometric specification, which expands on the mixed approach in the literature by including migration into adjacent occupations.

3.3.1 Ordering Occupations

Our paper tests the hypothesis that migrant inflow into occupations adjacent to occupation o, i.e. either *below* or *above*, affects native wages in occupation o. For this purpose, we rank occupations according to the order given by the SOC 2010 definition. Where the 1-digit SOC 2010 occupations are organised in order of 'skill level'. The SOC approximates skill levels using the "length of time deemed necessary for a person to become fully competent in the performance of the task associated with a job". This is determined by considering the formal training, qualifications and experience that may be required for the job [5]. We show our ranking in column (1) of table 3.2. As a robustness check, we also rank occupations by their real hourly wage, shown in Column 2 of Table 3.2 , as migrants tend to concentrate at the low end of the wage distribution (Dustmann et al., 2013).

To highlight our methodology, consider Professional occupations as an illustrative example. The occupation adjacent and above to *Professionals* are *Managers, Directors and Senior Officials* whereas the occupation adjacent and below to *Professionals* are *Associate Professionals and Technical Occupations*. Since managers are the highest and elementary the lowest occupations, we are dropping these occupations from our estimations.

Using occupations to define skill groups has the advantage that it allows us to avoid

[5]Office for National Statistics (2010), *Standard Occupational Classification 2010: Volume 1 - Structure and description of unit groups*, Palgrave Macmillan, ISBN 978-0-230-24819-9 Web Link

Table 3.2. Ranking of Occupations

Order	Rank	(1) **Standard ONS Order**	(2) **Real Hourly Wage Order**
Highest	1	Managers, Directors, Senior Officials	Managers, Directors, Senior Officials
	2	Professional Occupations	Professional Occupations
	3	Associate Professionals and Technical Occupations	Associate Professional and Technical Occupations
	4	Administrative and Secretarial Occupations	Skilled Trades Occupations
	5	Skilled Trades Occupations	Administrative and Secretarial Occupations
	6	Caring Leisure and Other Service Occupations	Process, Plant and Machine Operatives
	7	Sales and Customer Service Occupations	Caring Leisure and Other Service Occupations
	8	Process, Plant and Machine Operatives	Sales and Customer Service Occupations
Lowest	9	Elementary Occupations	Elementary Occupations

Notes: This table shows how we rank 9 1-digit SOC 2010 Occupations from Highest to Lowest. Column 1 ranks them by the standard ordering provided by and Column 2 is the ONS the average UK real hourly wage for each occupation.

the issue of migrant downgrading upon arrival in the UK, which is where migrants with high levels of education tend to work in jobs below that skill level. In such a case, parameter estimates would over-estimate the number of highly educated migrants competing with highly educated natives where in reality these highly educated migrants are also competing with lower educated natives. By using occupations to define skill groups we overcome the issue of downgrading and assume that managers compete with managers, professionals with professionals and so on.

3.3.2 Empirical Model

Our methodology builds upon the analysis by Dustmann et al. (2013), where the authors use UK data to estimate the total effect of migration into a region on native wages across the wage distribution within that region using the following specification

$$\triangle lnW^N_{prt} = \beta_p \triangle m_{rt} + \triangle X_{prt} + \gamma_t + \triangle\epsilon_{rt} \tag{3.1}$$

where $\triangle lnW^N_{prt}$ is the yearly change ($lnW^N_{prt} - lnW^N_{prt-1}$) in average log native wages in percentiles p, region r and time t and $\triangle m_{rt}$ is the yearly change in the migrant native ratio within a region r and time t. The migrant native ratio is defined as $m_{rt} = \frac{M_{rt}}{N_{rt}}$, i.e. the number of migrants working divided by the total number of natives in region r in year t. Moreover, the authors control for region characteristics X_{rt}, and time fixed effects, γ_t.

In order to estimate cross-occupational effects of migrants, we build off this model and divide each region-year observation into 9 occupations. Where our goal is to estimate whether the total effect of migration into the occupation below a native worker has a significant impact on that native worker's wages on average. Following the literature we first difference out any time invariant differences between regions and occupations. We then further control for any variation overtime for the UK as a whole by including time fixed effects. Our dependent variable becomes the yearly change in average log native wage, $\triangle lnW^N_{ort}$, in occupation group o in region r in year t. Using the occupational ranking outlined in the Ranking of Occupations section, we relate changes in native wages to three migration measures: i) yearly changes in the migrant-native ratio in the same occupational group o ($\triangle m_{ort}$, ii) yearly changes in the migrant-native ratio in the occupational group above o ($\triangle m_{o+1rt}$) iii) yearly changes in the migrant-native ratio in the occupational group below o ($\triangle m_{o-1rt}$) in region r and year t as follows

$$\triangle lnW^N_{ort} = \alpha + \beta_1 \triangle m_{ort} + \beta_2 \triangle m_{o+1rt} + \beta_3 \triangle m_{o-1rt} + \beta_4 \triangle X_{ort} + \gamma_t + \triangle \epsilon_{ort} \tag{3.2}$$

where X_{ort} denotes controls for the average age for natives and migrants and education controls, defined by the age they left education, for the proportion of migrants and natives with higher (25≥), high (20-24), intermediate (16-19) and low education (16 <) all within an occupation-region-time group and time fixed effects. The remaining variables are defined as above. We estimate robust standard errors clustered at the occupation specific regional level. One key issue when allowing for spatial variation is that it is possible for natives to react to migration, by for example moving to a different region. This would result in our coefficient being biased towards zero. We follow Dustmann et al. (2013) and use broad definitions of spatial regions which will reduce the likelihood of this being the case.[6]

A common concern when estimating the impact of migration on native wages is the endogenous allocation of migration into occupations and regions.
Results would be upwardly biased if migrants move to occupations and regions experiencing high growth. We use the standard approach in the literature and use a shift-share instrument. This instrument follows studies such as Bartel (1989) and Munshi (2003) which show that immigrants tend to migrate to where there is other migrants. Therefore, in the absence of shocks to the local labour market and occupations we would expect these new migrants to distribute themselves across the UK according to where migrants from their region are already settled in the UK. Following Card (2001) we construct an instrument which captures the 'supply-push' component of immigration inflows, that is those flows that are exogenous to local demand shocks and are the result of migrants moving to areas with other migrants similar to themselves. [7]. To construct this instrument, we divide migrants into 10 broad regions of origin where we would expect the

[6]Dustmann et al. (2013) test this using LFS data and find no evidence for a native response and in Chapter 3 I find that for 18 government office regions there is no evidence of a native response using the 2004-2019 APS.

[7]Jaeger et al. (2018) show that if the distribution of the country of origin of migrants within a region remains stable overtime then using a shift-share instrument may not be sufficient. As past migration shares are constructed using the 1991 Census then this is unlikely to be the case, where the Treaty of Maastricht introduced freedom of movement to the EU in 1992, which was then subsequently expanded to Austria, Finland and Sweden in 1995, and to ten Central and Eastern European countries in 2004 and further expanded in 2007. There were also increases in migration due to the UK handover of Hong Kong to China, and to a lesser extent an increase due to forced migration events. Furthermore, Borusyak et al. (2022) show that if these out-migration shocks from sending countries can be considered as-good-as-random then we are able to identify a causal effect, even when the initial migrant shares are not exogenous. In our setting, this is a plausible assumption. It is unlikely that the introduction of freedom of movement was selected based upon occupation-regional specific characteristics. This also applies to the subsequent expansion to freedom of movement. However, one concern in this setting, particularly with the EU expansions, is that this attracted migrants to mainly low skilled occupations, which could be experiencing different wage trends in this period compared to other occupations resulting in a correlation.

network effect to be stronger between migrants from similar regions [8]. We then calculate the change in the number of migrants from each region of origin, j, who enter the UK in year, t. The next step is to determine the exogenous distribution of migrants from each country of origin into each UK region. To ensure an exogenous regional distribution we must choose a sufficiently lagged period such that the regional share is not correlated with growth in the region, as such we use the regional shares from the 1991 Census 2% Sample of Anonymised Records, λ_{jr91}. Next, we calculate the occupation shares for migrants from each country of origin in year t, τ_{jot}. The exogenous inflow of immigrants with country of origin j with occupation o that is expected to move into labour market region r is therefore given by $\lambda_{jr91} \times \tau_{jot} \times \triangle M_{jt}$. To obtain the total exogenous inflow of migrants we then sum the inflows across all country of origins. Finally, to normalise our instrument we divide it by the overall occupation-specific labour force in region r which we lagged three times L_{ort-3}. By lagging our labour force variable we can account for any population changes induced by higher growth rates and avoid any correlation between the skill-specific error term. $\triangle\epsilon_{ort}$ in our main regression equation (Glitz, 2012). The advantage of using three lags over shorter lags is that it reduces the likelihood that the different population levels are correlated with higher growth rates. Where if wage growth is correlated overtime then it is possible that those groups with a higher population and wage growth in the previous one or two years also have higher population and wage growth today (Mitaritonna et al., 2017). This instrument is valid so long as past immigrant shares do not correlate with changes recent changes in economic growth within government office regions. Since we use the 1991 Census data this is unlikely to be the case.

$$SP_{jort} = \frac{\sum_j \lambda_{jr91}\tau_{jot}\triangle M_{jt}}{L_{ort-3}} \tag{3.3}$$

[8]Where we group migrants into 10 broad regions: Republic of Ireland, Old Commonwealth, Western Europe and Cyprus , Central Europe, Turkey and Former USSR , Africa, Indian Subcontinent , Caribbean and Other America , Middle East , Other Asia, Rest of the World

Unlike previous studies we must also instrument for the endogeneity of migration into below and above occupations. We use a Two Stage Leasr Squares (2SLS) approach where our first stage regresses the migrant-native ratio on our constructed supply-psuh instrument. This uses the same controls and time fixed effects outlined in equation 3.2. Where our first stage passes the weak instrument test with a Kleibergen-Paap F-stat of 24.99.

There are some potential issues with measurement error due to the sample being split into occupation, region, time groups. Where our sample size for migrants can be quite small in some groups, in for example regions like Northern Ireland, which can be exacerbated by first differencing our regression as pointed out by Dustmann et al. (2013). However, according to the authors using an instrumental variable estimation will account for this measurement error as long as the instrumental variable's measurement error is not correlated with the measurement error of our variable of interest. As we use the 1991 Census to construct the regional shares in our instrument, we do not expect this to be an issue. Finally, following Dustmann et al. (2013) we do not use the APS sample weights which are calculated for the whole population, and not migrants and natives separately.

3.4 Results

As a starting point we present the standard spatial results found in the UK migration literature. Thereafter, we estimate reduced form results for cross-occupational effects of migrants on wages for the whole of the UK and along the occupational distribution.

3.4.1 Regional Results

Before considering cross-occupational effects of migration, we show that we can replicate the spatial results of the paper that is the basis for our analysis Dustmann et al. (2013) pretty closely. In Table 3.3 we divide the UK into 13 regions and estimate the average effect of migration within a region for three time periods. The dependent variable is the change in the log native wage within a region time cell and we control for migrants and natives average age and the proportion of migrants and natives with low, intermediate, high and higher education within a region, time cell as well as for time fixed effects. Columns 1 and 2 present our OLS and Columns 3 and 4 our 2SLS results where we instrument the migrant-native ratio using the 1991 census. Columns 1 and 3 present results with time fixed effects but no extra controls and Columns 2 and 4 present results wiht both time fixed effects and extra controls. We find in our preferred specification in Column 4 that a 1 percentage point increase in migration increases real native wages by 0.638%. Although this is larger in size compared to Dustmann et al. (2013), which could be a result of the different sample period.

Table 3.3. The impact of migration on native wages: Spatial Regression

Dependent Variable	OLS		IV	
$\triangle$ Log Real Hourly Wages	(1)	(2)	(3)	(4)
$\triangle$ migration: own occupation	0.692***	0.802***	0.556***	0.638**
	(0.128)	(0.148)	(0.0655)	(0.235)
Observations	143	143	143	143
F-stat			2420.0	693.0
First-Stage Coefficient			1.634***	1.604***
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) migrants to native ratio $\triangle m_{rt}$ for the years 2004-2017. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. F-stat is the first stage Kleibergen-Paap F-stat testing for weak instruments. Clustered standard errors are reported in parentheses. $***p < 0.001$ $**p < 0.01$, $*p < 0.05$

3.4.2 Main Wage Results: Occupations ordered by standard ONS SOC 2010

The results shown in Table 3.4 are based on a panel dataset where for each year between 2004 and 2017 we divide the UK into 13 regions and each region again into 9 occupational groups. Although since the top and bottom occupation drop out, and we lag our instrument three times, our sample for analysis are 7 occupations between 2007-2017, resulting in 1001 occupation, region, time groups overall. As explained in section 3.3.1, for our main specification we rank these 9 occupations by the the standard SOC 2010 ordering provided by the ONS. The dependent variable is the change in the log real native hourly wage within an occupation, region, year group. Our controls are as follows: migrants and natives average age and the proportion of migrants and natives with low, intermediate, high and higher education within an occupation, region, year group. Following Dustmann et al. (2013) we also control for year fixed effects.

Columns 1 and 2 presents our OLS and Columns 3 and 4 our Two-Stage Least Squares (2SLS) results where our constructed Supply Push variable is the instrument. This instrument is valid as it has a strong and significant relationship with our main explanatory variable, where is passes the Kleibergen-Paap underidentification test at a 10% level of significance, and it passes the weak instrument test where the Kleibergen-Paap F-stat is over 10. Columns 1 and 3 present results with year fixed effects but no extra controls and Columns 2 and 4 present results with both year fixed effects and extra controls. If we consider Columns 2 and 4, there are two reasons that OLS coefficients may be different from IV coefficients in our setting. Firstly, there may be a positive bias in OLS, as a result of migrants moving into occupations with high wage growth. Secondly, there may be a bias towards zero as a result of an attenuation bias as a result of measurement error (Dustmann et al., 2013, Aydemir and Borjas, 2011). We expect that the former effect would be stronger for own occupations, as higher wage growth for occupation o would incentivise migration into occupation *o*. However, this is less likely to be the case for below and above occupations. For example, higher wage growth in occupation *o-1* is unlikely to incentivise migration into occupation *o*. In this scenario it may be the case that the attenuation bias is why OLS estimates are close to zero but become more positive for below occupations and more negative for above occupations.

Across all four models, the change in the migrant native ratio in the same region and occupation are insignificant. By contrast looking at our preferred model in Column 4, when we consider the migrant-native ratio in the occupation *below*, the 2SLS results suggest that a 1 percentage point increase in the migrant-native ratio in the occupation below a native's own occupation, within the same region and time, resulted in an increase in native wages of 0.332 percent. We can reject the hypothesis that the coefficients are statistically different from zero. For the migrant-native ratio in the occupations *above*, however, the results are negative, just under half the size of our below coefficient and we cannot reject the hypothesis that the coefficients are statistically different from zero. These results are smaller yet still comparable in size with Dustmann et al. (2013), whose results vary between 0.256-0.465 depending on the instrument used. Furthermore, they are not too dissimilar to the cross-effects from high school dropouts found by Borjas and Monras (2017) for the Mariel Boatlift which varies between 0.131-0.589.

This suggests that overall, the impact on native wages would be positive, where only migration into below occupations is significant. Furthermore, even if we consider the negative impact from above occupations we would expect the overall impact to remain positive. In section 3.2, Figures 3.4 and 3.5 shows that overtime there is a larger inflow of migrants into lower paid occupations compared to higher paid occupations. As such, we would expect the positive impact from below migration to be larger on average than any negative impact from higher paid occupations. More specifically, Appendix B.3 shows the average yearly percentage point change in the migrant-native ratio in the same, below and above occupations in our sample period. By multiplying this with our coefficient we find that the average yearly effect of migration into the same occupation for below occupations it is 0.329% and for above occupations it is -0.109%.

Table 3.4. Impact of migration on native wages: Standard ONS SOC 2010 Ordering

Dependent Variable	OLS		IV	
△ Log Real Hourly Wages	(1)	(2)	(3)	(4)
△ migration: own occupation	0.0741	0.0365	0.00931	-0.135
	(0.0941)	(0.0893)	(0.138)	(0.154)
△ migration:below occupation	0.157*	0.166*	0.312***	0.332**
	(0.0604)	(0.0668)	(0.0948)	(0.115)
△ migration: above occupation	0.00444	0.00543	-0.249	-0.136
	(0.143)	(0.151)	(0.176)	(0.171)
Observations	1001	1001	1001	1001
F-stat			37.69	24.99
Underidentification(p-value)			0.0690	0.0658
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in in the working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) employed migrants to native ratio $\triangle m_{ort}$ in the own, below and above occupations for the years 2004-2017. All estimations include 9 occupation groups ordered by standard ONS SOC 2010 ordering and are estimates using 13 government office regions. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. F-stat is the first stage Kleibergen-Paap F-stat testing for weak instruments. Clustered standard errors are reported in parentheses. $***p < 0.001$ $**p < 0.01$, $*p < 0.05$

3.4.3 Results along the occupation distribution

In Table 3.5 we investigate cross-occupational effects along the occupational distribution by splitting our sample in High and Low paid occupations. [9] As before, cross-occupational effects cannot be estimated for the highest and lowest occupations. The final groups, therefore, consist of 4 High paid occupations and 3 Low paid occupations. The results of Table 3.5 show that the positive effect of the migrant-native ratio for occupations below is concentrated in low paid occupations, but we cannot detect a significant result, likely due to the lower sample size. For low paid occupations, a 1 percentage point increase in the migrant-native ratio in an occupations below increases native wages by 0.316 percent, mirroring the effect for the whole sample quite closely. For migration into the occupation above however, the effect is positive and very close to zero.This may be due to highly educated migrants downgrading to work in lower paid occupation, resulting in a higher likelihood for productivity spillovers and positive peer effects on the native workforce, which we explore further in our mechanisms section. Overall, the reduced significance of the results could be due to the reduced sample size as a result of splitting the sample. The high paid occupation results are not as reliable as low paid, as it does not pass the weak instrument test and has much higher standard errors. However, it still shows a positive correlation in below occupations, although at a lower magnitude compared to low paid occupations.

[9]Low and High paid occupations are defined on whether they are above or below the median wage. Low Paid occupations include: Caring, Leisure and Other Services; Sales and Customer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations, and High Paid includes: Managers, Directors and Senior Officials; Professionals; Associate Professional and Technical; Administrative and Secretarial; Skilled Trades.

Table 3.5. Impact of migration on native wages in high and low paid occupations: Standard ONS SOC 2010 Ordering

Dependent Variable	OLS		IV	
△ Log Real Hourly Wages	(1)	(2)	(3)	(4)
Panel A: High Paid Occupations				
△ migration own occupation	0.442*	0.381	-0.0343	0.155
	(0.210)	(0.212)	(0.491)	(0.472)
△ migration below occupation	0.153	0.236	0.187	0.179
	(0.112)	(0.125)	(0.311)	(0.354)
△ migration above occupation	-0.0831	-0.203	-0.237	-0.479
	(0.235)	(0.281)	(0.698)	(0.993)
Observations	572	572	572	572
F-stat			2.637	1.472
Underidentification(p-value)			0.123	0.131
Panel B: Low Paid Occupations				
△ migration own occupation	-0.0365	-0.0641	0.105	-0.0957
	(0.0956)	(0.0733)	(0.134)	(0.279)
△ migration below occupation	0.143*	0.133	0.289*	0.316
	(0.0595)	(0.0781)	(0.114)	(0.192)
△ migration above occupation	0.0555	0.0676	-0.239	0.00614
	(0.187)	(0.158)	(0.170)	(0.206)
Observations	429	429	429	429
F-stat			19.06	21.68
Underidentification(p-value)			0.0877	0.0699
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) employed migrants to native ratio $\triangle m_{ort}$ in the own, below and above in high and low occupations for the years 2004-2017. All estimations include 9 occupation groups ordered by SOC 2010 where High Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median and are estimates using 13 government office regions. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. F-stat is the first stage Kleibergen-Paap F-stat testing for weak instruments. Clustered standard errors are reported in parentheses.$***p < 0.001$ $**p < 0.01$, $*p < 0.05$

3.5 Robustness Checks

A key assumption we have made throughout the paper is that the best way to order occupations is the standard SOC 2010 occupation ordering provided by the ONS. To test the robustness of our results, we test an alternative way to order occupations, by ordering them by their average real hourly wages.

3.5.1 Wage Results: Ordered by Real Hourly Wage

The results shown in Table 3.6 use the same panel dataset and estimation strategy as Table 4. As explained in section 3.3.1, as a robustness check we rank these 9 occupations by the mean real hourly wage of their employees.

Columns 1 and 2 present our OLS and Columns 3 and 4 our Two-Stage Least Squares (2SLS) results where our constructed Supply Push variable is the instrument. This instrument is valid as it has a strong and significant relationship with our main explanatory variable,where is passes the Kleibergen-Paap underidentification test at a 5% level of significance, and it passes the weak instrument and underidentification test where the Kleibergen-Paap F-stat is over 10. Columns 1 and 3 present results with year fixed effects but no extra controls and Columns 2 and 4 present results with both year fixed effects and extra controls. Across all four models, the change in the migrant-native ratio in the same region and occupation are positive but insignificant. By contrast looking at our preferred model in Column 4, when we consider the migrant-native ratio in the occupation *below*, the 2SLS results suggest that a 1 percent increase in the change in the migrant native ratio in the occupation below a native's own occupation, within the same region and time, resulted in a statistically significant increase of 0.251 percent.For the migrant-native ratio in the occupations *above*, however, the results are the opposite. Where a 1 percent increase in the migrant native ratio in the occupation above a native's own occupations, within the same region and time, resulted in an statistically significant increase of -.209 percent.

These results differ slightly than our main results. Where although below occupations are positive and significant the magnitude is slightly smaller than our standard SOC 2010 ordering. Furthermore, in this specification our magnitude for migration into above occupations is still negative but of a larger magnitude and significant, suggesting the results for migration into above occupations are not as robust for migration into below occupations. Nevertheless, it is still the case that the overall cross-effects of migration are positive. Replicating our analysis of the average yearly cross-effects of migration, we multiply the average yearly percentage point change in the migrant-native ratio in above and below occupations from 2007-2017 with the coefficients obtained in Table 3.6. We find migration into below occupations effects native wages by 0.245% a year on average, whereas above occupations only effect native wages by -0.17% on average.

3.5.2 Results along the occupational distribution: Ordered by Real Hourly Wage

In Table 3.7 we investigate the robustness of our cross-occupational effects along the occupational distribution by splitting our sample in High and Low paid occupations for the real hourly wage ordering. This does not change the occupations in high and paid occupations but it does change the ordering of occupations within each group. [10] As before, cross-occupational effects cannot be estimated for the highest and lowest occupations. The final groups, therefore, consist of 4 High paid occupations and 3 Low paid

[10] Low and High paid occupations are defined on whether they are above or below the median wage. Low Paid occupations include: Caring, Leisure and Other Services; Sales and Customer Service Occupations; Process, Plant and Machine Operatives; Elementary Occupations, and High Paid includes: Managers, Directors and Senior Officials; Professionals; Associate Professional and Technical; Administrative and Secretarial; Skilled Trades.

Table 3.6. Impact of migration on native wages: Ordered by Real Hourly Wage

Dependent Variable	OLS		2SLS	
△ Log Real Hourly Wages	(1)	(2)	(3)	(4)
△ migration: own occupation	0.130	0.0908	0.166	0.0812
	(0.0853)	(0.0753)	(0.127)	(0.120)
△ migration: below occupation	-0.0425	-0.0353	0.246***	0.251***
	(0.0772)	(0.0755)	(0.0565)	(0.0528)
△ migration: above occupation	0.0598	0.0796	-0.276***	-0.209**
	(0.0576)	(0.0748)	(0.0743)	(0.0710)
Occupation-Region-Year Groups	1001	1001	1001	1001
First Stage F-stat			11.13	10.88
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Entries are estimated regression coefficients of the yearly change in in the working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) employed migrants to native ratio $\triangle m_{ort}$ in the own, below and above occupations for the years 2004-2017. All estimations include 9 occupation groups ordered by real hourly wages and are estimates using 13 government office regions. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. F-stat is the first stage Kleibergen-Paap F-stat testing for weak instruments. Clustered standard errors are reported in parentheses. $*p < 0.001 * *p < 0.01, *p < 0.05$

occupations. Like our standard occupation ordering results results, Table 3.7 shows that the positive effect of the migrant-native ratio for occupations below is concentrated in low paid occupations, finding even stronger effects compared to our main specification in Table 3.5. For low paid occupations, a 1 percent increase in the migrant-native ratio in occupations below increases native wages by 0.253 percent which is almost exactly the same as the effect for the whole sample. For migration into the occupation above however, the effect is still negative but are insignificant, this could be due to the reduced sample size by splitting the sample. These results further add to the evidence that due to downgrading of highly educated migrants into lower paid occupations, there is a higher likelihood for productivity spillovers and positive peer effects on the native workforce, which we explore in the next section. The results for high paid occupations still face a weak instrument problem and are not as reliable as the results for low paid occupations.

Table 3.7. Impact of migration on native wages in high and low paid occupations: Ordered by Real Hourly Wage

Dependent Variable	OLS		IV	
△ Log Real Hourly Wages	(1)	(2)	(3)	(4)
Panel A: High Paid Occupations				
△ migration: own occupation	0.458*	0.396	0.0114	0.0477
	(0.206)	(0.198)	(0.137)	(0.153)
△ migration: below occupation	-0.0310	-0.0135	0.564	0.645
	(0.0895)	(0.0801)	(0.404)	(0.436)
△ migration: above occupation	0.219	0.216	-0.629	-0.738
	(0.166)	(0.134)	(0.627)	(0.678)
Occupation-Region-Year Groups	572	572	572	572
F-stat			3.359	3.318
Panel B: Low Paid Occupations				
△ migration: own occupation	0.0320	-0.00404	0.250*	0.160
	(0.104)	(0.0810)	(0.122)	(0.161)
△ migration: below occupation	-0.0577	-0.0593	0.203***	0.253***
	(0.119)	(0.127)	(0.0573)	(0.0721)
△ migration: above occupation	0.00929	0.0684	-0.287***	-0.150
	(0.0726)	(0.134)	(0.0637)	(0.121)
Occupation-Region-Year Groups	429	429	429	429
F-stat			10.17	10.49
Year Dummies	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in working age(16-64) employed native log real hourly wages on the yearly change in the working age(16-64) employed migrants to native ratio $\triangle m_{ort}$ in the own, below and above in high and low occupations for the years 2004-2017. All estimations include 9 occupation groups ordered by real hourly wages where High Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median and are estimates using 13 government office regions. Additional covariates are controls for migrants and natives separately and include average age, the proportion with higher, high, intermediate and low education, and year fixed effects. F-stat is the first stage Kleibergen-Paap F-stat testing for weak instruments. Clustered standard errors are reported in parentheses. $*p < 0.001 * *p < 0.01, *p < 0.05$

3.6 Potential Mechanisms

After presenting the reduced form cross-occupational effects, we explore the role of two potential mechanisms through which the results may operate. First, we concentrate on productivity changes and second we investigate whether migrant inflow into occupations below can allow natives to specialise into better renumerated tasks.

3.6.1 Productivity

As a consequence of migrants downgrading, relatively higher educated migrants take up employment in low paying occupations, especially just after arriving in the UK. This inflow of highly educated migrants into a region-occupation cell likely increases the average level of education in that particular cell. As a result of this, natives who work with higher educated migrants may be subject to stronger peer effects, for example as a result of larger knowledge spillovers, which would increase productivity and therefore wages. If we assume natives interact not only with migrants in adjacent occupations, or if their productivity benefits from more productive workers adjacent to theirs, then we may expect that this would result in positive spillovers across occupations also. Furthermore, an increase in the average education in an adjacent skill cell could also result in productivity gains between complementary occupations by increasing the ability of workers in each occupation to more effectively work together. To investigate this, we firstly estimate how the change in the migrant-native ratio, m_{ort} correlates with the change in the proportion of the total sample with high or higher education, within occupation, region, time groups. We define an individual as having high or higher education, $\triangle Educ^{H}_{ort}$ and regress yearly changes of this variable on changes in the migrant-native ratio in that particular cell, $\triangle m_{ort}$, in a specification similar to the one outlines in Equation 1. We also control for changes in the average age and the proportion with low education, the proportion with intermediate education for both migrants and natives separately. Table 3.8 documents a strong, consistent and statistically significant correlation between the migrant-native ratio in an occupation and the proportion of highly educated individuals in that occupations. For the whole sample, an increase in the migrant-native ratio is correlated with an increase in the overall proportion with high and higher education across all occupations at 0.0862. We then run the same regression, splitting occupations into high and low paid occupations as before. We find the correlation between migrants and educational attainment is similar along the occupational distribution. As a next step, we estimate whether wages in an occupation relate to changes in the proportion of highly educated native workers in occupations above and below. For this we regress changes in native log real wages $\triangle lnW^{N}_{ort}$ on the changes in the proportion of all workers in the same occupation, $\triangle Educ^{H}_{ort}$, the occupation above, $\triangle Educ^{H}_{o+1rt}$, and the occupation below $\triangle Educ^{H}_{o-1rt}$, who have high or higher education in a regression. We use a regression framework similar to the one outlined in equation 1 and include controls for age and the average proportion of low and intermediate workers for migrants and natives in the same occupation, region, time group and age controls and time dummies. Unsurprisingly, column 1 table 3.9 shows that the proportion of highly educated individuals working in occupation o is positively associated with mean wages in occupation o, however this association is not statistically significant. More relevant to our purposes however, we find a positive but much weaker correlation changes in educational attainments of employees in an occupation below occupation o and changes in native wages in occupation o. The results for above occupations are robust to how occupations are ordered as shown in Appendix B.4. Although many of these results are insignificant, it may be useful to consider why the cross-occupational correlation from above tends to be negative but is positive or zero from below. One explanation is that there are stronger peer effects from below than from above. For example, a hard-worker in an occupation below you encourages you to also work harder, but if someone being paid more than you works harder then this is to be expected and may in fact encourage you to shirk if you know your slack will be covered. Alternatively, it could be the case that any productivity gains from complementarity with more educated migrant workers tend to be allocated to those in already higher paying occupations. Such that you gain more compared to those below you, but less than those above you. Our results along the wage distribution suggest that we should see stronger productivity cross-effects in low paid occupations compared to high paid occupations. Column 3 shows this is the case, where low paid occupations see a weak positive correlation in both below and above occupations but in high paid occupations this is very close to zero. The results for low paid occupations are however not robust, where in Appendix B.4 we order our occupations by real hourly wages and find a negative correlation for migration into occupations below, and a positive correlation for above. Although these results are much closer to zero. Furthermore, we find a larger significantly negative correlation above for high paid occupations, this coincides with the significant negative impact found in our robustness checks and adds further suggestive evidence that those above you may discourage productivity or at the very least take most productivity gains from complementarity with those in occupations they are one rank above.

Table 3.8. Correlations between migration and education

Dependent Variable	OLS	
△ High Education	(1)	(2)
All Occupations		
△ migration: own occupation	0.0636**	0.0862**
	(0.0209)	(0.0265)
Observations	1287	1287
High Paid Occupations		
△ migration: own occupation	0.0624	0.108***
	(0.0434)	(0.0223)
Observations	715	715
Low Paid Occupations		
△ migration own: occupation	0.0659*	0.0803**
	(0.0255)	(0.0294)
Observations	572	572
Year Dummies	Yes	Yes
Other Controls	No	Yes

Notes: Entries are estimated regression coefficients of the yearly change in the overall proportion with high and higher education on the yearly change in the employed migrant to native ratio, $\triangle m_{ort}$ in the own, below and above occupations for All, High and Low SOC occupations for years 2007-2017, where ordering does not matter for this regression. Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median and are estimated using 13 government office regions. Additional covariates are controls for migrants and natives separately and include average age, the proportion with intermediate and low education and year fixed effects. Clustered standard errors are reported in parentheses. $***p < 0.01, **\ p < 0.05 * P < 0.05$

Table 3.9. Correlation between education changes and wages: Standard ONS SOC 2010 Ordering

Dependent Variable △ Log Real Hourly Wage	(1) All	(2) High	(3) Low
△ Native Low	-0.307	-0.788	0.0821
Educated own occ	(0.253)	(0.811)	(0.251)
△ Migrant Low	0.108	0.112	0.0701
Educated own occ	(0.0768)	(0.0712)	(0.154)
△ Native Intermediate	0.0405	0.187	0.0454
Educated own occ	(0.212)	(0.560)	(0.234)
△ Migrant Intermediate	-0.0417	-0.114	0.0237
Educated own occ	(0.0391)	(0.0757)	(0.0371)
△ High Educated own occ	0.0821	0.261	-0.0942
	(0.308)	(0.789)	(0.327)
△ High Educated below occ	0.00188	0.00699	0.0253
	(0.00826)	(0.00778)	(0.0808)
△ High Educated above occ	-0.00943*	-0.00561	0.0118
	(0.00534)	(0.00742)	(0.0759)
Observations	1001	572	429
Controls	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Notes: Entries are estimated regression coefficients of the yearly change in native log real hourly wages on the yearly change in the overall proportion with high and higher education in the own, below and above occupations for All, High, and Low SOC 2010 occupations using the standard ONS ordering for years 2007-2017. Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median and are estimated using 13 government office regions.. Additional covariates are controls for migrants and natives separately and include the average age, the proportion with intermediate and low education and year fixed effects. Clustered standard errors are reported in parentheses.
$***p < 0.001$ $**p < 0.01$, $*p < 0.05$

3.6.2 Specialisation

We also consider an alternative mechanism of impact where migrant inflow into lower occupations allows natives to specialise in more complex, better renumerated tasks. As previously highlighted by Peri and Sparber (2009) migrant inflows can allow natives to specialise in jobs which are more concentrated in complex communicative and interactive tasks, for which they have a comparative advantage. This could occur from natives moving into occupations which focus on these tasks, alternatively, they natives in an adjacent occupation could specialise in more communicative tasks as a response. This specialisation could result in an increase in overall productivity and therefore an increase in native wages.

One channel through which we could see cross-effects would be from natives moving to occupations with more complex tasks. However, we lack the necessary variables to map each occupation to a particular task and so are unable to test for this outcome. Alternatively, it could also be from natives in an adjacent occupation changing their task specialization as a response. To illustrate this, consider two four-digit occupations, one from the Professional Occupations – Civil Engineers, and one from Associate Professionals – Engineering Technicians. The latter may be responsible for more manual tasks such as setting up equipment, performing calculations, recording and interpreting data. Whereas a Civil Engineer may be more responsible for the organization and design of these projects. An increase in migration into the Engineering Technicians occupation could reduce the time Civil Engineers may spend helping with the more manual tasks, resulting in an evolution of their roles that may incentivise Civil Engineers to specialize in the more communicative tasks associated with the role.

Our ability to test this directly is still limited by the variables available to us, but in this scenario we can take an alternative approach. We approximate specialisation into more technical tasks by the proportion of natives either taking up or being offered 'Job Related Training or Education'. Since specialising into more technical tasks is likely to entail some re-training, the offer or completion of job related training might approximate natives specialising.

We define two averages, one for the proportion of natives undertaking 'Job Related Training or Education' and one for the proportion of natives being offered 'Job Related Training or Education'. We then regress yearly changes in these proportions on yearly changes in the migrant-native ratio in the same occupations ($\triangle m_{ort}$), in the occupation below ($\triangle m_{ort}$), in the occupation above ($\triangle m_{ort}$) within a region, time cell. We use a framework analogous to equation 1 where we also control for the total average proportion of higher, high, intermediate and low educated native and migrant workers in the same occupation, region, time group and age controls and time dummies. Table 3.10 column 2 shows a weak but nonetheless detectable correlation between changes in the migrant-native ratio in the same occupation and changes in native employees undertaking 'Job Related Training or Education'. However, the correlates are insignificant for migration into above and below. Appendix B.4 table B.7 shows that when we order occupations by real hourly wages then the results are similar.

This shows that there is little evidence of migration into an occupation resulting in increased task specialisation for natives in adjacent occupations, however there is suggestive evidence this specialisation may occur within broad occupation groups. Future studies should consider whether this is the result of natives moving occupations within the broad occupation groups, or whether it is the task content of their current occupation changing. In addition, a more careful causal analysis of specialising between groups would have to be undertaken to gain a better picture of this mechanism.

Table 3.10. Correlation between migration and native training - Standard ONS SOC 2010 ordering

	Training/Education Completed		Training/Education Offered	
	(1)	(2)	(3)	(4)
	No Control	Control	No Control	Control
$\triangle$ migration:	0.0781^{**}	0.0815^{**}	0.0799	0.0748
own occupation	(0.0383)	(0.0378)	(0.0500)	(0.0512)
$\triangle$ migration:	0.00409	0.00276	0.00479	0.00284
below occupation	(0.0234)	(0.0245)	(0.0351)	(0.0361)
$\triangle$ migration:	-0.0419	-0.0388	0.0213	0.0284
above occupation	(0.0546)	(0.0536)	(0.0560)	(0.0578)
Observations	1001	1001	1001	1001
Time FE	Yes	Yes	Yes	Yes
Degrees of Freedom	13	23	13	23

Notes: Entries are estimated regression coefficients of the yearly change in the proporiton of natives who have taken or have been offered but rejected job related training or education on the yearly change in the employed migrant to native ratio, $\triangle m_{ort}$ in the own, below and above SOC 2010 occupations using the standard ONS ordering for years 2007-2017. All estimations include 9 occupation groups ordered by real hourly wages and are estmated using 13 government office regions. Additional covariates are controls for migrants and natives separately and include the average age, the proportion with higher, high, intermediate and low education and year fixed effects. Clustered standard errors are reported in parentheses. $***p < 0.001$ $**p < 0.01, *p < 0.05$

3.7 Conclusion

The results presented in this paper suggest that the wages of natives working in an occupation and a region are increased by immigration into lower paying occupations into the same region. This effect is strongest for low paying occupations, which tallies with results presented by Dustmann et al. (2013) showing that migrants to the UK downgrade upon arrival, although due to reduced sample size the results should be interpreted with caution. We identify and test two likely mechanisms. First, we find that immigration into an occupation increases the average educational attainment of all employees working in that occupation. This effect arises mechanically by immigrants being more educated than natives. The average educational attainment of an occupation in the same and below occupation, in turn, is weakly but positively associated with wages in higher paying occupations, however we find a weak negative association for above occupations. Second, we find that immigration into an occupation increases in-job training offers of natives working in the same occupation, and a weaker relationship for natives in a better paid occupations, which possibly allows natives to specialise into better remunerated tasks. Our results have important implications for policy makers. Much of the policy debate surrounding migration focuses on how to attract high skilled migrants for high skilled jobs. Our results, however, suggest policymakers should consider the wider work environment and the complementarities that can occur across occupations. If countries stop migration into low skilled occupations then this could potentially reduce productivity spillovers to natives in higher paid occupations and thus harm real wage growth for natives, which in the UK has remained noticeably low since the financial crisis. Future studies would benefit from a more in-depth and causal exploration of potential mechanisms to better understand where these spillovers arise from.

Bibliography

D. Acemoglu, T. A. Hassan, and J. A. Robinson. Social structure and development: A legacy of the Holocaust in Russia. *The Quarterly Journal of Economics*, 126(2):895–946, 2011. 23, 45, 72

M. Akbulut-Yuksel and M. Yuksel. The long-term direct and external effects of Jewish expulsions in Nazi Germany. *American Economic Journal: Economic Policy*, 7(3):58–85, 2015. 24

M. Alfano, R. Mckenzie, and G. Roy. Cross-occupational effects of immigration on native wages in the uk. *Strathclyde Discussion Papers in Economics*, 2020. 111

M. Alsan. The effect of the tsetse fly on African development. *American Economic Review*, 105(1):382–410, 2015. 36

O. Altındağ, O. Bakış, and S. V. Rozo. Blessing or burden? impacts of refugees on businesses and the informal economy. *Journal of Development Economics*, 146:102490, 2020. 21

J. G. Altonji and D. Card. The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market*, pages 201–234. University of Chicago Press, 1991. 82, 91, 93, 96, 116

B. Apor. *The expulsion of the 'German' communities from Eastern Europe at the end of the Second*

World War, chapter The expulsion of the German speaking population from Hungary, pages 33–47. European University Institute, 2004. 21, 25, 29, 30, 31, 34, 38

C. E. Arbatli and G. Gokmen. Ethnic capital and sub-national development: Armenian and Greek legacy in post-expulsion Turkey. *Available at SSRN 3273964*, 2018. 23, 50, 72

A. Aydemir and G. Borjas. Cross-country variation in the impact of international migration: Canada, mexico and the united states. *Journal of the European Economic Association*, 5(4):663–708, 2007. 95, 115

A. Aydemir and G. J. Borjas. Attenuation bias in measuring the wage impact of immigration. *Journal of Labor Economics*, 29(1):69–112, 2011. 138

P. Azoulay, J. S. Graff Zivin, and J. Wang. Superstar extinction. *The Quarterly Journal of Economics*, 125(2):549–589, 2010. 116

B. Balassa. Reforming the new economic mechanism in hungary. *Journal of Comparative Economics*, 7(3):253–276, 1983. 47, 56

B. Balkan and S. Tumen. Immigration and prices: quasi-experimental evidence from Syrian refugees in turkey. *Journal of Population Economics*, 29(3):657–686, 2016. 21

A. P. Bartel. Where do the new us immigrants live? *Journal of Labor Economics*, 7(4): 371–391, 1989. 98, 133

G. Basso and G. Peri. The association between immigration and labor market outcomes in the united states. 2015. 82, 92, 96

S. O. Becker and A. Ferrara. Consequences of forced migration: A survey of recent findings. *Labour Economics*, 59:1–16, 2019. 21

S. O. Becker, K. Boeckh, C. Hainz, and L. Woessmann. The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, 126(590):40–74, 2016. 24, 36, 66

S. O. Becker, I. Grosfeld, P. Grosjean, N. Voigtlander, and E. Zhuravskaya. Forced migration and human capital: Evidence from post-WWII population transfers. *American Economic Review*, 110(5):1430–63, 2020. 24

S. O. Becker, J. Rubin, and L. Woessmann. Religion in economic history: A survey. *The Handbook of Historical Economics*, pages 585–639, 2021. 69

P. Bharadwaj, A. I. Khwaja, and A. Mian. Population exchange and its impact on literacy, occupation and gender–evidence from the partition of India. *International Migration*, 53 (4):90–106, 2015. 24

M. Bisello. How does immigration affect natives' task-specialisation? evidence from the united kingdom. ISER Working Paper Series 2014-12, University of Essex, Colchester, 2014. URL `http://hdl.handle.net/10419/95670`. 117

H. Bleakley and J. Lin. Portage and path dependence. *The Quarterly Journal of Economics*, 127(2):587–644, 2012. 36

D. Borbely and R. Mckenzie. Forced migration and local economic development: Evidence from postwar hungary. *Strathclyde Discussion Papers in Economics*, 2021. 20

G. J. Borjas. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118(4):1335–1374, 2003. 89, 90, 95, 115

G. J. Borjas. Native internal migration and the labour market impact of immigration. *Journalof Human Resources*, 41(2):221–258, 2006. 80, 91, 95, 106, 116

G. J. Borjas. *Immigration economics*. Harvard University, 2014. 88, 90, 95, 115

G. J. Borjas and J. Monras. The labour market consequences of refugee supply shocks. *Economic Policy*, 32(91):361–413, 2017. 139

K. Borusyak, P. Hull, and X. Jaravel. Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213, 2022. 133

L. P. Boustan, P. V. Fishback, and S. Kantor. The effect of internal migration on local labor markets: American cities during the great depression. *Journal of Labor Economics*, 28(4): 719–746, 2010. 91, 96

O. R. M. S. P. Bratsberg, Bernt; Raaum. Immigration wage effect by origin. *The Scandinavian Journal of Economics*, 116(2):356–393, 2014. 95

S. Braun and M. Kvasnicka. Immigration and structural change: Evidence from post-war Germany. *Journal of International Economics*, 93(2):253–269, 2014. 24

R. Breunig, N. Deutscher, and H. T. To. The relationship between immigration to australia and the labour market outcomes of australian-born workers. *Economic Record*, 93(301): 255–276, 2017. 91, 95, 96

D. L. Brown and K. A. Schafft. Population deconcentration in Hungary during the post-socialist transformation. *Journal of Rural Studies*, 18(3):233–244, 2002. 38, 46, 47

B. Callaway, A. Goodman-Bacon, and P. H. Sant'Anna. Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*, 2021. 74, 75

F. Campo, G. Forte, and J. Portes. The impact of migration on productivity and native-born workers' training. ., 2018. 115, 116

D. Card. The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2): 245–257, 1990. 82, 116

D. Card. Using regional variation in wages to measure the effects of the federal minimum wage. *Ilr Review*, 46(1):22–37, 1992. 44, 74

D. Card. Immigrant inflows, native outflows and the local market impacts of higher migration. *Journal of Labor Economics*, 19(1):22–64, 2001. 8, 9, 80, 91, 95, 98, 113, 116, 133

D. Card. How immigration affects u.s. cities. *CReAM Discussion Paper 711*, 2007. 91, 93, 96

D. Card. Immigration and inequality. *American Economic Review: Papers and Proceedings.*, 99(2):1–21, 2009. 91, 95, 116

D. Card and E. G. Lewis. "the diffusion of mexican immigrants during the 1990s: Explanations and impacts.". In G. J. Borjas, editor, *Mexican immigration to the United States*, chapter 6, pages 193–228. University of Chicago Press, 2007. 91, 95

D. Card and G. Peri. Immigration economics by george j. borjas: A review essay. *Journal of Economic Literature*, 54(4):1333–49, 2016. 7, 8, 79, 80, 81, 82, 87, 88, 89, 90, 91, 92, 93, 95, 102, 103, 107, 109, 115

E. Ceritoglu, H. B. G. Yunculer, H. Torun, and S. Tumen. The impact of Syrian refugees on natives' labor market outcomes in Turkey: evidence from a quasi-experimental design. *IZA Journal of Labor Policy*, 6(1):1–28, 2017. 21

E. Chaney and R. Hornbeck. Economic dynamics in the Malthusian era: evidence from the 1609 Spanish expulsion of the Moriscos. *The Economic Journal*, 126(594):1404–1440, 2016. 24

Y.-w. Chen, L. Ni, and L. Ospina-Forero. Visualising internal migration flows across local authorities in england and wales. *Environment and Planning A: Economy and Space*, 53 (4):616–618, 2021. 98

M. A. Clemens, E. G. Lewis, and H. M. Postel. Immigration restrictions as active labor market policy: Evidence from the mexican bracero exclusion. *American Economic Review*, 108(6):1468–87, 2018a. 116

M. A. Clemens, E. G. Lewis, and H. M. Postel. Immigration restrictions as active labor market policy: Evidence from the Mexican Bracero exclusion. *American Economic Review*, 108(6):1468–87, 2018b. 72

T. Cornelissen, C. Dustmann, and U. Schönberg. Peer effects in the workplace. *American Economic Review*, 107(2):425–56, 2017. 114, 116

D. S. Cornelius. *Hungary in World War II: Caught in the cauldron*. Fordham Univ Press, 2011. 29

F. D'Amuri and G. Peri. Immigration, jobs, and employment protection: evidence from europe before and during the great recession. *Journal of the European Economic Association*, 12(2):432–464, 2014. 117

D. R. Davis and D. E. Weinstein. Bones, bombs, and break points: the geography of economic activity. *American Economic Review*, 92(5):1269–1289, 2002. 36

C. Dustmann and A. Glitz. How do industries and firms respond to changes in local labor supply? *Journal of Labour Economics*, 33(3):711–750, 2015. 95, 116

C. Dustmann, F. Fabbri, and I. Preston. The impact of immigration on the british labour market. *The Economic Journal*, 115(507):F324–F341, 2005. 82, 89, 92, 96

C. Dustmann, T. Frattini, and I. Preston. The effect of immigration along the distribution of wages. *Review of Economic Studies*, 80(1):145–173, 2013. 81, 82, 84, 89, 91, 92, 96, 97, 98, 100, 106, 112, 113, 114, 116, 127, 130, 131, 133, 135, 136, 138, 139, 154

C. Dustmann, U. Schonberg, and J. Stuhler. The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, 30(4):31–56, 2016. 8, 9, 79, 82, 89, 90, 112, 115

C. Dustmann, U. Schönberg, and J. Stuhler. Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics*, 132(1):435–483, 2017. 80, 92, 96, 116

A. Edo. The impact of immigration on the labor market. *Journal of Economic Surveys*, 33(3): 922–948, 2019. 8, 82

A. Edo. The impact of immigration on wage dynamics: Evidence from the algerian independence war. *Journal of the European Economic Association*, 18(6):3210–3260, 2020. 81, 82, 93, 96, 109, 116

P. Elias, M. Birch, et al. Soc2010: revision of the standard occupational classification. *Economic & Labour Market Review*, 4(7):48–55, 2010. 189

R. J. W. Evans. *Austria, Hungary, and the Habsburgs: Central Europe C. 1683-1867*. Oxford University Press, 2006. 25

A. Falk and A. Ichino. Clean evidence on peer effects. *Journal of labor economics*, 24(1): 39–57, 2006. 116

H. Fischer. *Räumliche Aspekte des sozio-ökonomischen Wandels der ungarndeutschen Minderheit im 20. Jahrhundert*. na, 1992. 21, 27

M. Foged and G. Peri. Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, 8(2):1–34, 2016. 82, 91, 96, 116, 117

B. Gatti. Emlékezet, örökség, identitás: bukovinai székelyek néprajzi, kulturális antropológiai vizsgálata. 2019. 70

M. Gehrsitz and M. Ungerer. Jobs, crime, and votes: A short-run evaluation of the refugee crisis in Germany. 2017. 21

A. Glitz. The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in germany. *Journal of Labor Economics*, 30(1):175–213, 2012. 91, 95, 116, 134

M. Goos and A. Manning. Lousy and lovely jobs: The rising polarization of work in britain. *The review of economics and statistics*, 89(1):118–133, 2007. 119, 189

D. M. Gould. Immigrant links to the home country: empirical implications for us bilateral trade flows. *The Review of Economics and Statistics*, pages 302–316, 1994. 116

I. Grosfeld, A. Rodnyansky, and E. Zhuravskaya. Persistent antimarket culture: a legacy of the pale of settlement after the Holocaust. *American Economic Journal: Economic Policy*, 5(3):189–226, 2013. 24

P. Grosjean. The weight of history on European cultural integration: A gravity approach. *American Economic Review*, 101(3):504–08, 2011. 24, 36

A. Grosz. Keresztutakon (*Crossroads*). *ARGUMENTUM Publishing*, 2020(1):65, 2020. 69

H. Inalcik. *The Ottoman Empire: 1300-1600*. Hachette UK, 2013. 26

C. K. Jackson and E. Bruegmann. Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4): 85–108, 2009. 116

D. A. Jaeger, J. Ruist, and J. Stuhler. Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research, 2018. 82, 92, 96, 98, 116, 133

T. Kemeny and A. Cooke. Spillovers from immigrant diversity in cities. *Journal of Economic Geography*, 18(1):213–245, 2018. 116

W. R. Kerr and W. F. Lincoln. The supply side of innovation: H-1b visa reforms and us ethnic invention. *Journal of Labor Economics*, 28(3):473–508, 2010. 116

R. J. LaLonde and R. H. Topel. Labor market adjustments to increased immigration. In *Immigration, Trade and the Labor Market*, pages 167–199. University of Chicago Press, 1991. 95, 116

J. Lee, G. Peri, and V. Yasenov. The employment effects of Mexican repatriations: Evidence from the 1930's. Technical report, National Bureau of Economic Research, 2017. 39

J. Lee, G. Peri, and V. Yasenov. The labor market effects of mexican repatriations: Longitudinal evidence from the 1930s. Technical report, National Bureau of Economic Research, 2019. 116

S. Lemos and J. Portes. New labour? the effects of migration from central and eastern europe on unemployment and wages in the uk. *The BE Journal of Economic Analysis & Policy*, 14(1):299–338, 2013. 81, 82, 92, 96, 116

E. Lewis. Immigration, skill mix, and capital skill complementarity. *The Quarterly Journal of Economics*, 126(2):1029–1069, 2011. 91, 95, 116

J. Llull. The effect of immigration on wages: Exploiting exogenous variation at the national level. Technical report, Centre for Research and Analysis of Migration(CReAM), Department of Economics, University College London., 2014. 95

J. Llull. The effect of immigration on wages: exploiting exogenous variation at the national level. *Journal of Human Resources*, pages 0315–7032R2, 2017. 95, 115

P. R. Magocsi. *Historical atlas of central Europe*. University of Toronto Press, 2002. 26

R. Marchut. *Töréspontok: A Budapest környéki németség második világháborút követő felelősségre vonása és annak előzményei (1920–1948)*. MTA Társadalomtudományi Kutatóközpont, 2014. 21, 22, 27, 32, 35, 51, 64, 66, 67, 68, 69, 72

A. Mas and E. Moretti. Peers at work. *American Economic Review*, 99(1):112–45, 2009. 116

C. Mitaritonna, G. Orefice, and G. Peri. Immigrants and firms' outcomes: Evidence from france. *European Economic Review*, 96:62–82, 2017. 82, 91, 93, 96, 99, 116, 134

K. Munshi. Networks in the modern economy: Mexican migrants in the us labor market. *The Quarterly Journal of Economics*, 118(2):549–599, 2003. 98, 133

S. Nickell and J. Saleheen. The impact of immigration on occupational wages: evidence from britain. Staff Working Paper 575, Bank of England, 2015. Staff Working Paper. 116

S. Nickell and J. Saleheen. The impact of eu and non-eu immigration on british wages. *IZA Journal of Development and Migration*, 7(1):1–28, 2017. 89, 91, 92, 95, 116

N. Nunn and L. Wantchekon. The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–52, 2011. 24, 66

J. L. M. Olea and C. Pflueger. A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369, 2013. 99

F. Ortega and G. Peri. Openness and income: The roles of trade and migration. *Journal of international Economics*, 92(2):231–251, 2014. 95, 115, 116

J. Ortega and G. Verdugo. Who stays and who leaves? immigration and the selection of natives across locations. *Journal of Economic Geography*, 22(2):221–260, 2022. 80, 82, 93, 96, 116

G. I. Ottaviano and G. Peri. The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography*, 6(1):9–44, 2006. 116

G. I. Ottaviano, G. Peri, and G. C. Wright. Immigration, offshoring, and american jobs. *American Economic Review*, 103(5):1925–59, 2013. 116

G. I. Ottaviano, G. Peri, and G. C. Wright. Immigration, trade and productivity in services: Evidence from uk firms. *Journal of International Economics*, 112:88–108, 2018. 112, 116

C. Ozden and M. Wagner. Immigrants vs natives? displacement and job creation. Unpublished Paper, 2016. 95

L. Pascali. Banks and development: Jewish communities in the Italian Renaissance and current economic performance. *Review of Economics and Statistics*, 98(1):140–158, 2016. 24

G. Peri. The effect of immigration on productivity: Evidence from us states. *Review of Economics and Statistics*, 94(1):348–358, 2012. 117

G. Peri and C. Sparber. Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3):135–69, 2009. 10, 112, 115, 152

G. Peri and V. Yasenov. The labor market effects of a refugee wave synthetic control method meets the mariel boatlift. *Journal of Human Resources*, 54(2):267–309, 2019. 82, 116

G. Peri, K. Shih, and C. Sparber. Stem workers, h-1b visas, and productivity in us cities. *Journal of Labor Economics*, 33(S1):S225–S255, 2015. 112, 116

H. Rolfe, C. Rienzo, M. Lalani, and J. Portes. Migration and productivity: employers' practices, public attitudes and statistical evidence. *National Institute of Economic and Social Research*, 2013. 116

F. Sá. Immigration and house prices in the uk. *The Economic Journal*, 125(587):1393–1424, 2015. 81, 82, 92, 96, 106, 107, 116

A. Salvatori. The anatomy of job polarisation in the uk. *Journal for labour market research*, 52(1):8, 2018. 119, 122, 190

A. Schumann. Persistence of population shocks: Evidence from the occupation of West Germany after world war ii. *American Economic Journal: Applied Economics*, 6(3):189–205, 2014. 24

G. Seewann. Geschichte der deutschen in ungarn. 2012. 21, 25, 26, 27, 31, 67

J. Segnana, J. J. Endara, et al. Short-term effects of forced displacement on host communities: evidence from the Rohingya crisis. Technical report, Asociación Argentina de Economía Política, 2020. 21

A. Semrad. Immigration and educational spillovers: evidence from Sudeten German expellees in post-war Bavaria. Technical report, Munich Discussion Paper, 2015. 24

J. Sharpe and C. R. Bollinger. Who competes with whom? using occupation characteristics to estimate the impact of immigration on native wages. *Labour Economics*, 66:101902, 2020. 90, 95, 115

W. R. Shepherd. *Historical atlas*. H. Holt, 1911. 26

T. Spira. Worlds apart: The Swabian expulsion from Hungary after World War ii. *Nationalities Papers*, 13(2):188–197, 1985. 21, 27

M. Steinhardt. The wage impact of immigration in germany - new evidence for skill groups and occupations. *The B.E. Journal of Economic Analysis & Policy*, 11(1):1–35, 2011. 95

P. A. Testa. The economic legacy of expulsion: Lessons from Post-war Czechoslovakia. *The Economic Journal*, 131(637):2233–2271, 2021. 24

A. Toth. Települések magyarországon 1945–1948 között. *A németek kitelepítése, a belső népmozgások és a szlovák–magyar lakosságcsere összefüggései*, 1993. 21, 22, 25, 30, 31, 32, 33, 35, 51, 52, 63, 64, 66

A. Toth. Integráció elméletben-diszkrimináció a gyakorlatban: németek magyarországon, 1948-1956. *PRO MINORITATE*, 2016:47–55, 2016. 32

A. Toth. *Dokumentumok a magyarországi németek történetéhez 1944-1953= Quellen zur Geschichte der Deutschen in Ungarn 1944–1953*. Argumentum Kiadó, 2018. 29

S. Tumen. The economic impact of Syrian refugees on host countries: Quasi-experimental evidence from Turkey. *American Economic Review*, 106(5):456–60, 2016. 21

H.-J. Voth. Persistence–myth and mystery. In *The Handbook of Historical Economics*, pages 243–267. Elsevier, 2021. 22, 45

F. Waldinger. Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *The Review of Economic Studies*, 79(2):838–861, 2012. 116

N. J. White. Internal migration, england and wales: Year ending june 2015. Technical report, Office of National Statistics, 2017. 98

R. Winter-Ebmer and J. Zweimüller. Immigration and the earnings of young native workers. *Oxford economic papers*, 48(3):473–491, 1996. 96

Appendix A

Appendix - Chapter 1

A.1 Additional Tables and Figures

Table A.1. OLS regression of Share of Germans in 1941 on Ottoman Occupation Dummy

	(1) Share of Germans in 1941	(2) Share of Germans in 1941
Formerly Ottoman Occupied Counties	0.113*** (0.008)	0.112*** (0.008)
Observations	3032	2872
Controls	No	Yes

Notes: This table presents estimates from an OLS regression of the share of Germans in each township in 1941 on a dummy indicating Ottoman occupation for counties whose current population centre were under Ottoman occupied territory furing the 16th and 17th centuries. Column (2) includes township-level covariates such as level of urbanisation and population size. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.2. Sample Selection - OLS Regression of the Share of Germans in 1941 on a dummy for townships dropped from the sample

	(1) 1920-1990 Census Sample	(2) 1949 Census Sample	(3) 2011 Census Sample
Township Not in Final Sample	-0.038 (0.024)	0.004 (0.008)	-0.003 (0.008)
Observations	3032	3032	3032

Notes: This table presents estimates from an OLS regression of the share of Germans in each township in 1941 on a dummy indicating whether a township was dropped from our three main samples due to not having data on outcomes or forced migration intensity for years after 1941. The sample used for this analysis is the cross-section of townships present in our 1941 Census sample. Column (1) shows results for the sample where these data are matched with the 1920-1990 sample of townships for whom population density data in provided through the 1990 Census. Column (2) shows results for the sample where 1949 Census data are matched in. Column (3) shows results for the sample where 2011 Census data are matched in. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3. Sample Selection - Selected vs Non-Selected Sample

	1920-90 Census Sample			1949 Census Sample			2011 Census Sample		
	Selected	Non-selected	*p-value*	Selected	Non-selected	*p-value*	Selected	Non-selected	*p-value*
Covariates (1941)									
Urban Area	0.05	0.21	0.02	0.03	0.06	0.00	0.05	0.06	0.20
Area in 1941 (km2)	3009.01	17721.20	0.24	2416.32	3494.26	0.00	2851.75	3672.72	0.11
Population (1941)	2422.14	46613.14	0.27	1788.38	3463.79	0.03	2258.02	4500.49	0.17
Nr of townships	2991	41		833	2199		1973	1059	

Notes: This table show results for the comparison of means for the full 1941 Census sample of townships and our three main analytical samples: the 1920-90 Census sample that we use to examine the effects of forced migration on population density (Columns 1-3); the 1949 Census sample that we use to examine the effects of forced migration on various short-term outcomes (Columns 4-6); and the 2011 Census sample that we use to examine the effects of forced migration on various long-term outcomes (Columns 7-9). Means for the selected sample and the non-selected sample are reported in Columns 1,2,4,5,7, and 8. Columns 3,6, and 9 report the p-value for the statistical test of the mean differences.

Table A.4. Summary Statistics - 1949 Census Sample

	Mean	SD	Count
Treatment			
Share of Germans in 1941	0.06	0.20	833
Forced Migration Intensity (Registry Data)	0.03	0.13	833
1941 Census			
Population (1941)	1921.38	4313.40	833
Population Density (1941)	0.73	0.52	833
Share of Agricultural Workers (1941)	0.33	0.15	819
Share of Trade Workers (1941)	0.01	0.01	730
Share of Transport Workers (1941)	0.01	0.01	766
Share of Manufacturing Workers (1941)	0.03	0.03	812
Share Employed (1941)	0.45	0.08	830
1949 Census			
Population (1949)	1788.38	2824.81	799
Population Density (1949)	0.72	0.40	799
Share of Agricultural Workers (1949)	0.11	0.19	833
Share of Trade Workers (1949)	0.02	0.07	833
Share of Manufacturing Workers (1949)	0.18	0.62	833
Labour Force Density (1949)	0.37	1.19	833
Share Employed (1949)	0.45	0.16	799

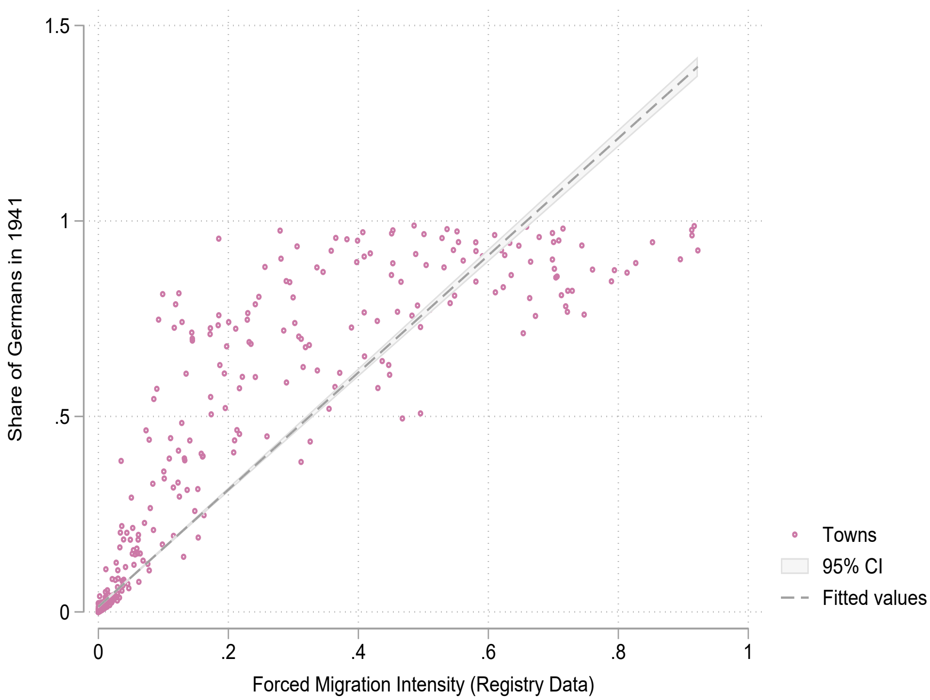
Notes: These summary statistics are based on the sample of townships where we could merge Census data from 1941 with the same data in 1949. This is the sample used for our baseline analysis of short-term effects.

Table A.5. Summary Statistics - 2011 Census Sample

	Mean	SD	Count
Treatment			
Share of Germans in 1941	0.07	0.21	1973
Forced Migration Intensity (Registry Data)	0.04	0.13	1973
Covariates (1941)			
Share Employed (1941)	0.43	0.10	1549
Population (1941)	2783.72	7180.57	1973
Share of Physical Workers (1941)	0.14	0.08	1524
Share of Helpers (1941)	0.12	0.06	1537
Share of Services Workers (1941)	0.01	0.01	1401
Share of Trade Workers (1941)	0.01	0.01	1361
Share of Transport Workers (1941)	0.01	0.01	1429
Share of Agricultural Workers (1941)	0.32	0.15	1530
Share of Manufacturing Workers (1941)	0.02	0.03	1520
Share of Intellectuals (1941)	0.02	0.06	1521
Covariates			
Non-Arrable Land Area (1000 hectares)	191.95	361.27	1973
Arrable Land Area (1000 hectares)	5685.77	10142.08	1973
Area Suitable for Cultivation of Main Crops (1000 hectares)	1798.33	5206.35	1971
Distance from Austrian Border	417.35	108.41	1973
Distance from Eastern Border	323.93	123.37	1973
Outcomes			
Agriculture Share of Labour	0.07	0.06	1762
Manufacturing Share of Labour	0.36	0.10	1929
Trade Share of Labour	0.15	0.04	1856
Labour Force Density	0.24	0.45	1973
Share Employed	0.35	0.08	1973

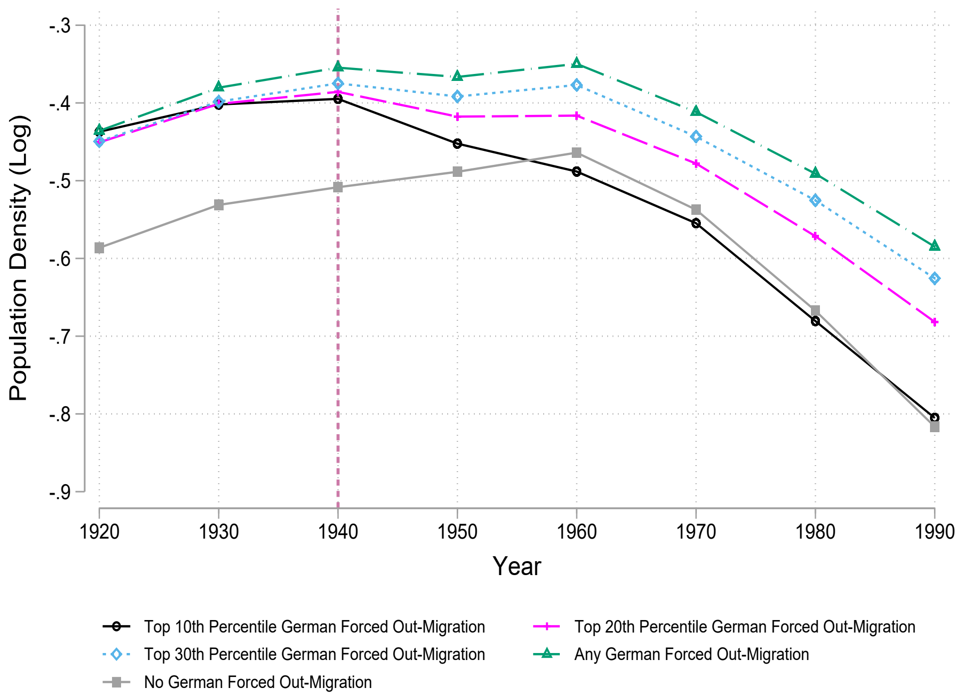
Notes: These summary statistics are based on the sample of townships where we could merge Census data from 1941 with the same data in 2011. This is the sample used for our baseline analysis of persistent (contemporary) effects.

Figure A.1. The Share of Germans in 1941 and Forced Migration Intensity



Notes: The measure of forced migration intensity used for this figure is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. The dashed line represents fitted values from an OLS regression of the share of Germans in 1941 on forced migration intensity.

Figure A.2. Event Studies - Population Density Over Time



(a) Population Density (Log) Over Time

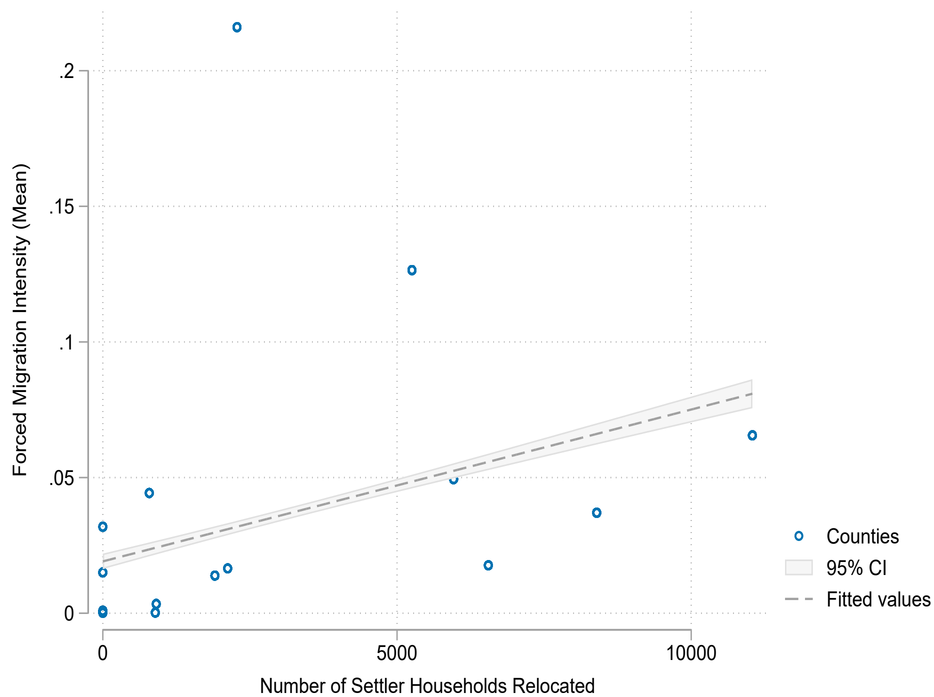
Notes: The measure of forced migration intensity is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. The black connected line with the circles plots (log) population density over time for the top 10th percentile forced migration townships. The grey connected line with the squares shows the same outcome for townships where no forced migration had taken place. The magenta long dashed line with plus signs is the top 20th percentile of forced migration townships. The dotted blue line with diamonds is the top 30th percentile of forced migration townships. The dot-dash green line with triangles is townships with any forced migration.

Table A.6. OLS Results - Population Density in 2011

Panel A:	Population Density (Log)		
	(1)	(2)	(3)
Share of Germans in 1941	-0.259***	-0.159*	-0.236**
	(0.077)	(0.091)	(0.095)
Number of Townships	1270	1270	1255
R^2	0.576	0.652	0.741
Panel B:	Population Density (Log)		
	(1)	(2)	(3)
Forced Migration Intensity (Census Data)	-0.274***	-0.172*	-0.247***
	(0.079)	(0.092)	(0.095)
Number of Townships	1270	1270	1255
R^2	0.576	0.653	0.742
Panel C:	Population Density (Log)		
	(1)	(2)	(3)
Forced Migration Intensity (Registry Data)	-0.388***	-0.189	-0.311**
	(0.131)	(0.143)	(0.141)
Number of Townships	1270	1270	1255
R^2	0.575	0.652	0.741
Covariates	No	Yes	Yes
County FE	Yes	Yes	Yes
Covariates (1941)	Yes	Yes	Yes
Area FE	No	No	Yes

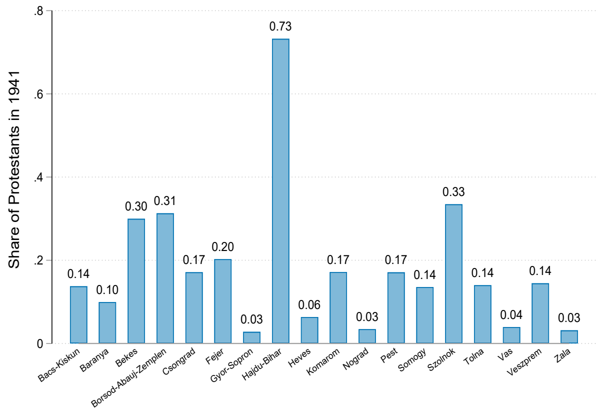
Notes: The point estimates are obtained from regressing the (log) population density in 2011 on our three forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data, and Panel C uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates from both 1941 and 2011. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Figure A.3. Hungarian Settler Inflows into Counties and Forced Migration Intensity

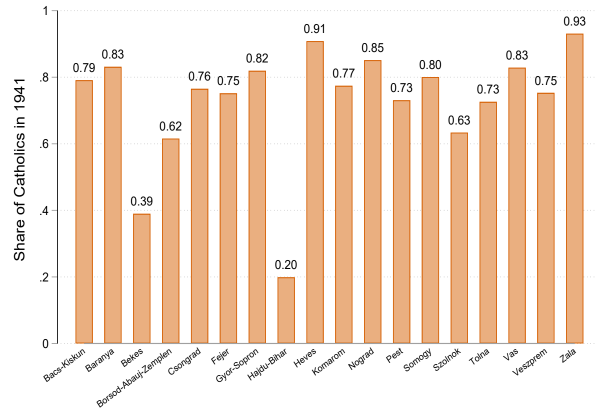


Notes: The measure of forced migration intensity used for this figure is based on the deportations registry data compiled by Hungarian authorities in 1946-1947. The dashed line represents fitted values from an OLS regression of the number of Hungarian settlers moving to each county in 1946 on forced migration intensity.

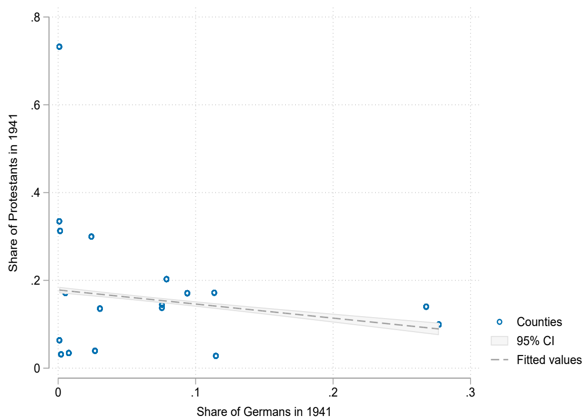
Figure A.4. Religious Composition of Hungarian Counties - 1940



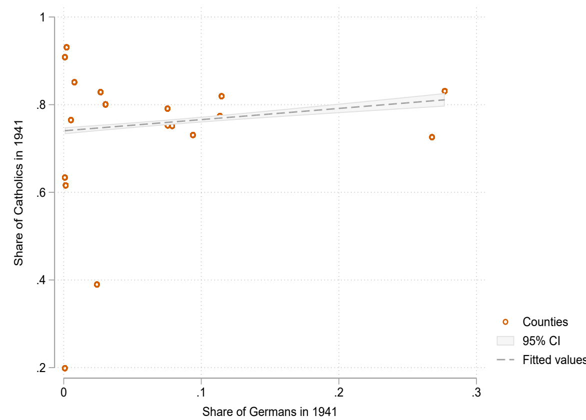
(a) Share of Protestants in 1941



(b) Share of Catholics in 1941



(c) Share of Protestants and Share of Germans in 1941



(d) Share of Catholics and Share of Germans in 1941

Notes: Data for this figure is obtained from the 1941 Census.

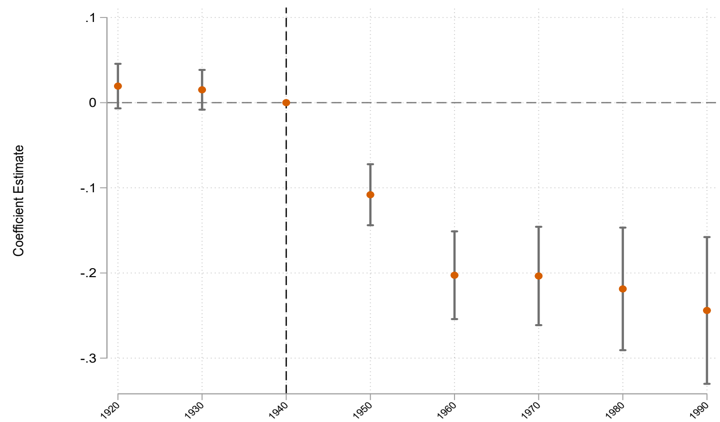
A.2 Tables and Figures for Robustness Checks

Table A.7. OLS Results - Population Density Over Time - Alternative Measures of Forced Migration Intensity

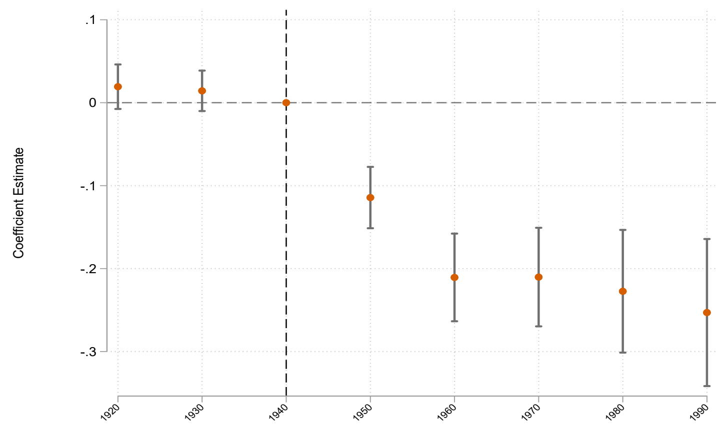
Panel A:	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Share of Germans in 1941 X Post	-0.115***	-0.139***	-0.110***	-0.100***	-0.079***
	(0.018)	(0.018)	(0.016)	(0.017)	(0.016)
Observations	8779	11626	14520	17398	20257
R^2	0.970	0.947	0.933	0.935	0.942
Mean DV	-0.48	-0.47	-0.47	-0.49	-0.52
SD DV	0.51	0.52	0.54	0.57	0.60
Panel B:	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Forced Migration Intensity (Census) X Post	-0.121***	-0.145***	-0.115***	-0.104***	-0.082***
	(0.019)	(0.018)	(0.017)	(0.018)	(0.016)
Observations	8770	11614	14505	17380	20236
R^2	0.970	0.947	0.933	0.935	0.942
Mean DV	-0.48	-0.47	-0.47	-0.49	-0.52
SD DV	0.51	0.52	0.54	0.57	0.60
Town FE	Yes	Yes	Yes	Yes	Yes
County x Year FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: The point estimates are obtained from regressing the log of population density on our two alternative measures of forced migration intensity. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, and Panel B uses our population adjusted measure based on Census data Each column shows the effect of the forced migrations in a different sample year. In each column specification, all preceding years are included in the sample, while all subsequent years are excluded. All specifications include county fixed effects, county times year fixed effects, and all our covariates. We control for 1941 population density in all specifications. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Figure A.5. Event Studies - Population Density Over Time - Alternative Forced Migration Intensity Measures



(a) Treatment: Share of Germans in 1940



(b) Treatment: Forced Migration Intensity (Census Data)

Notes: For panels a) and b), the point estimates plotted are from a version of our baseline regression where year fixed effects are interacted with the forced migration intensity variable. The baseline year is 1940. Confidence intervals (vertical bars) not spanning zero indicate significance at the 5% level.

Table A.8. OLS Results - Employment Rate in 1949 - Alternative Measures of Forced Migration Intensity

Panel A:	Employment Rate		
	(1)	(2)	(3)
Share of Germans in 1941	-0.059**	-0.080**	-0.044
	(0.029)	(0.031)	(0.036)
Number of Townships	799	664	664
R^2	0.006	0.096	0.128
Panel B:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Census Data)	-0.062**	-0.085***	-0.049
	(0.029)	(0.032)	(0.037)
Number of Townships	799	664	664
R^2	0.006	0.096	0.128
Covariates	No	Yes	Yes
County FE	No	No	Yes

Notes: The point estimates are obtained from regressing the employment rate in 1949 on our alternative forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data. The most demanding specification in Column 3 includes county fixed effects and all our covariates. Covariates include the 1941 labour market shares of different sectors and pre-migration employment rates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.9. OLS Results - Employment Rate in 2011 - Alternative Measures of Forced Migration Intensity

Panel A:	Employment Rate		
	(1)	(2)	(3)
Share of Germans in 1941	0.007	0.016	0.010
	(0.011)	(0.011)	(0.012)
Number of Townships	1270	1270	1255
R^2	0.385	0.413	0.545
Panel B:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Census Data)	0.007	0.016	0.010
	(0.011)	(0.011)	(0.012)
Number of Townships	1270	1270	1255
R^2	0.385	0.413	0.545
Covariates	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Labour Shares in 1941	No	Yes	Yes
Area FE	No	No	Yes

Notes: The point estimates are obtained from regressing the employment rate in 2011 on our alternative forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data. The most demanding specification in Column 3 includes county fixed effects, area fixed effects, and all our covariates. Covariates include the 1941 labour market shares of different sectors and pre-migration employment rates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.10. OLS Results - Population Density - Matched Sample

Panel A:	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Share of Germans in 1941 X Post	-0.096***	-0.113***	-0.079***	-0.079***	-0.078***
	(0.020)	(0.017)	(0.014)	(0.012)	(0.011)
Observations	2970	3934	4914	5888	6855
R^2	0.943	0.938	0.926	0.921	0.936
Mean DV	-0.37	-0.36	-0.37	-0.39	-0.42
SD DV	0.51	0.53	0.56	0.60	0.63
Panel B:	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Forced Migration Intensity (Census) X Post	-0.103***	-0.117***	-0.081***	-0.082***	-0.080***
	(0.021)	(0.017)	(0.014)	(0.013)	(0.012)
Observations	2970	3934	4914	5888	6855
R^2	0.943	0.938	0.926	0.921	0.936
Mean DV	-0.37	-0.36	-0.37	-0.39	-0.42
SD DV	0.51	0.53	0.56	0.60	0.64
Panel C:	Population Density (Log)				
	(1) 1950	(2) 1960	(3) 1970	(4) 1980	(5) 1990
Forced Migration Intensity X Post	-0.194***	-0.177***	-0.114***	-0.102***	-0.093***
	(0.033)	(0.029)	(0.021)	(0.021)	(0.020)
Observations	2970	3934	4914	5888	6855
R^2	0.944	0.938	0.926	0.921	0.936
Mean DV	-0.37	-0.36	-0.37	-0.39	-0.42
SD DV	0.51	0.53	0.56	0.60	0.64
Town FE	Yes	Yes	Yes	Yes	Yes
County x Year FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: This specification uses a sample of townships that was matched on observable characteristics through a propensity score matching (PSM) method. The point estimates are obtained from regressing the log of population density on our three forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data, and Panel C uses the population adjusted measure relying on Registry data. Each column shows the effect of the forced migrations in a different sample year. In each column specification, all preceding years are included in the sample, while all subsequent years are excluded. All specifications include town fixed effects, year fixed effects and county times year fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.11. OLS Results - Employment Rate in 1949 - Matched Sample

Panel A:	Employment Rate		
	(1)	(2)	(3)
Share of Germans in 1941	-0.068**	-0.069**	-0.062
	(0.033)	(0.034)	(0.041)
Observations	218	218	218
R^2	0.018	0.096	0.179
Panel B:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Census Data)	-0.072**	-0.073**	-0.067
	(0.033)	(0.035)	(0.042)
Observations	218	218	218
R^2	0.018	0.096	0.180
Panel C:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Registry Data)	-0.121**	-0.128**	-0.112
	(0.055)	(0.059)	(0.073)
Observations	218	218	218
R^2	0.021	0.100	0.180
Mean DV	0.42	0.42	0.42
SD DV	0.16	0.16	0.16
Covariates	No	Yes	Yes
County FE	No	No	Yes

Notes: This specification uses a sample of townships that was matched on observable characteristics through a propensity score matching (PSM) method. The point estimates are obtained from regressing the employment rate in 1949 on our three forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data, and Panel C uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates from 1941. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.12. OLS Results - Employment Rate in 2011 - Matched Sample

Panel A:	Employment Rate		
	(1)	(2)	(3)
Share of Germans in 1941	0.004	0.005	0.005
	(0.010)	(0.012)	(0.012)
Number of Townships	445	445	445
R^2	0.128	0.361	0.361
Panel B:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Census Data)	0.004	0.004	0.004
	(0.011)	(0.013)	(0.013)
Number of Townships	445	445	445
R^2	0.128	0.360	0.360
Panel C:	Employment Rate		
	(1)	(2)	(3)
Forced Migration Intensity (Registry Data)	-0.003	0.001	0.001
	(0.018)	(0.023)	(0.023)
Number of Townships	445	445	445
R^2	0.128	0.360	0.360
Covariates	Yes	Yes	Yes
County FE	No	Yes	Yes
Additional Controls	No	No	Yes
Area FE	No	No	Yes

Notes: This specification uses a sample of townships that was matched on observable characteristics through a propensity score matching (PSM) method. The point estimates are obtained from regressing the employment rate in 2011 on our three forced migration intensity measures. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data, and Panel C uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates from 1941. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Table A.13. OLS Results - Labour Share in 2011 - Matched Sample

Panel A:	Manufacturing Share of Labour			Agriculture Share of Labour			Trade Share of Labour			Labour Force Density		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Share of Germans in 1941	0.031***	0.009	0.009	0.034**	0.049***	0.049***	-0.015**	-0.011*	-0.011*	-0.007	-0.017	-0.017
	(0.010)	(0.011)	(0.011)	(0.014)	(0.015)	(0.015)	(0.006)	(0.007)	(0.007)	(0.053)	(0.057)	(0.057)
Number of Townships	410	410	410	440	440	440	423	423	423	445	445	445
R^2	0.150	0.266	0.266	0.095	0.277	0.277	0.097	0.202	0.202	0.293	0.390	0.390
Panel B:	Manufacturing Share of Labour			Agriculture Share of Labour			Trade Share of Labour			Labour Force Density		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Forced Migration Intensity (Census Data)	0.033***	0.011	0.011	0.032**	0.048***	0.048***	-0.015**	-0.012*	-0.012*	-0.014	-0.026	-0.026
	(0.011)	(0.012)	(0.012)	(0.015)	(0.016)	(0.016)	(0.006)	(0.007)	(0.007)	(0.053)	(0.058)	(0.058)
Number of Townships	410	410	410	440	440	440	423	423	423	445	445	445
R^2	0.152	0.266	0.266	0.094	0.275	0.275	0.097	0.201	0.201	0.293	0.390	0.390
Panel C:	Manufacturing Share of Labour			Agriculture Share of Labour			Trade Share of Labour			Labour Force Density		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Forced Migration Intensity (Registry Data)	0.045***	0.010	0.010	0.039*	0.069***	0.069***	-0.022**	-0.020*	-0.020*	-0.040	-0.060	-0.060
	(0.015)	(0.017)	(0.017)	(0.024)	(0.026)	(0.026)	(0.009)	(0.010)	(0.010)	(0.065)	(0.085)	(0.085)
Observations	410	410	410	440	440	440	423	423	423	445	445	445
R^2	0.144	0.265	0.265	0.090	0.272	0.272	0.095	0.203	0.203	0.293	0.390	0.390
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Additional Controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This specification uses a sample of townships that was matched on observable characteristics through a propensity score matching (PSM) method. The point estimates are obtained from regressing our outcomes (measured in 2011) on our three forced migration intensity measures. The outcomes are the manufacturing share of labour, the agriculture share of labour, the trade share of labour, and labour force density in each township. Panel A shows the results from the estimations using the share of Germans in 1941 as our measure of forced migration intensity, Panel B uses our population adjusted measure based on Census data, and Panel C uses the population adjusted measure relying on Registry data. The most demanding specification in Column 3 includes county fixed effects, additional controls for geographic variables, area ('jaras') fixed effects, and all our township level covariates. Covariates include 1941 labour market shares of different sectors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the township level.

Appendix B

Appendix - Chapter 2

B.1 Data Sampling

The APS utilises the LFS, LLFS and the APS(B) data however these have different designs which influence the construction of the APS. The LFS uses a rotational sampling design, such that a household will be in the sample for five quarters, where each quarter a cohort will drop out, one cohort will be on their last interview in wave 5 and one cohort will newly join the survey and be on wave 1. Information is collected 1-4 weeks prior to the interview, however questions on gross weekly wages and hours worked are only asked during the first and fifth interview. From the LFS the APS only utilises those who are either in their first interview(wave 1) and their last interview(wave 5). So within one year of the APS we will have 8 different sample groups from the LFS, two of which will be sampled each quarter which prevents the same household being included twice in one 4-quarter period. So this means between two consecutive years 50% of the sample will be in common. The LLFS sample is designed differently. Where households sampled will be interviewed in four annual waves, so the same household is interviewed four years in a row with the fieldwork spread equally across the four quarters. As such for each consecutive year 75% of the LLFS sample is in common and 25% is replaced. The LLFS sample is stratified by local area and the sample size is determined by a target number of Economically Active interviews. However if that target is achieved from wave 1 and 5 from the main LFS then no boost is required. If we consider the 2014 data, 319,757 responding or imputed people are from 155,554 households. 42.1% came from the main LFS and the rest from the LLFS(ONS). The APS(B) data was a sample boost for England only in the years 2004 and 2005. This sample did not answer all of the sample questions and as such some estimates from the APS are based on a subset of the database.

B.2 Bridging Occupations

The APS data obtains two different SOC definitions, shown in Table B.1. From 2004-2010 each observation is assigned 1-digit, 2-digit and 3-digit SOC 2000 code, and from 2011-2017 each observation is assigned a 1-digit, 2-digit and 3-digit SOC 2000 and a SOC 2010 code. There were four main areas of change from the SOC 2000- SOC 2010 codes. Firstly, managers were more strictly defined. Where, jobs with the manager title whose tasks did not involve significant responsibilities for strategic control over resources were reallocated to other major occupation groups. Secondly, there was a reallocation of most nursing occupations from associate professionals in group 3 and technical occupations to professional occupations in group 2. Thirdly, there was a reclassification of occupations associated with information technologies however this did not impact their allocation across major occupation groups. Lastly, there was a creation of supervisory unit groups at the 4-digit level in major occupation groups 4, 5, 6 and 7. It does not seem this directly impacted potential movements between major occupations but would indirectly do so through the stricter definition of managers in major group 1 (Elias et al., 2010). Following Goos and Manning (2007), we use unconditional matching to assign observations from 2004-2010 a 1-digit SOC 2010 occupations. Where we use the dual-coded 2011 APS dataset to estimate the proportion of each 3-digit SOC 2000 occupations that would be assigned into each 1-digit SOC 2010 occupations. Observations from 2004-2010 are then randomly assigned a 1-digit SOC 2010 definition such that the distribution replicates that found in the 2011 dataset. The potential downside to this approach is that we assume the reallocation is random and not conditional on for example gender or education Salvatori (2018).

Table B.2 shows that the characteristics of each 1-digit SOC occupation is very similar in terms of wages, age, migrants, education and female. Furthermore, Tables B.3 and B.4 show that the differences between the SOC 2010 characteristics in 2010 and 2011 are similarly different when comparing SOC 2000 characteristics from 2010 and 2011.

Table B.1. Definition of occupations

	(1)	(2)
Code	**SOC 2000**	**SOC 2010**
1	Managers and Senior Officials	Managers, Directors and Senior Officials
2	Professional Occ.	Professional Occ.
3	Associate Professional and Technical Occ.	Associate Professional and Technical Occ.
4	Administrative and Secretarial Occ.	Administrative and Secretarial Occ.
5	Skilled Trades Occ.	Skilled Trades Occ.
6	Personal Service Occ.	Caring, Leisure and Other Service Occ.
7	Sales and Customer Service Occ.	Sales and Customer Service Occ.
8	Process, Plant and Machine Operatives	Process, Plant and Machine Operatives
9	Elementary Occ.	Elementary Occ.

Notes: This table shows the definition of 1-digit SOC 2000 and 2010 occupations.

Table B.2. Summary Stats for SOC 2010 and SOC 2000 Occupations from 2004-2010

	1		2		3		4		5		6	
	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean
Real Hourly Wage	20.79	21.42	20.46	21.30	17.25	16.06	11.70	11.26	11.83	11.70	9.12	9.09
Age	43.67	43.34	42.48	42.53	40.47	40.51	41.32	41.28	41.18	41.15	40.18	40.08
Female	0.35	0.36	0.50	0.45	0.44	0.52	0.80	0.81	0.09	0.09	0.84	0.85
Migrants	0.09	0.09	0.12	0.12	0.09	0.10	0.07	0.07	0.07	0.07	0.09	0.09
Education												
Higher(>25)	0.02	0.02	0.07	0.08	0.03	0.03	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.26	0.28	0.49	0.56	0.30	0.30	0.14	0.14	0.08	0.07	0.10	0.10
Intermediate (19-21)	0.61	0.61	0.40	0.32	0.60	0.61	0.73	0.74	0.71	0.71	0.71	0.71
Low (16-18)	0.10	0.09	0.04	0.03	0.07	0.06	0.10	0.10	0.20	0.20	0.16	0.16
None/In Education	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02

	7		8		9	
	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean	soc10 Mean	soc2k Mean
Real Hourly Wage	8.88	8.13	10.66	10.56	8.02	8.02
Age	35.38	34.73	43.25	43.29	38.54	38.60
Female	0.69	0.71	0.14	0.14	0.48	0.48
Migrants	0.08	0.08	0.11	0.11	0.12	0.12
Education						
Higher(>25)	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.10	0.09	0.05	0.05	0.06	0.06
Intermediate (19-21)	0.63	0.62	0.68	0.68	0.62	0.62
Low (16-18)	0.13	0.13	0.26	0.26	0.22	0.22
None/In Education	0.14	0.16	0.01	0.01	0.10	0.10

Notes: Entries are for working age(16-64) natives and immigrants for the average real hourl wage, average age, share of female and the share in each education group. Higher educatoin: left full-time education after age 25, High education: left full-time education between age 20-24, Low education: left full-time education between age 16-19, None/Still in education: left education at age 15 or below, or is still in education. Occupation number follows the code assigned in Table B.1. Source: APS 2004-2011

Table B.3. Summary Stats for SOC2010 Occuations in 2010 and 2011

	1		2		3		4		5		6	
	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean
Real Hourly Wage	21.12	22.25	20.79	20.07	17.37	16.23	11.75	11.18	11.86	11.36	9.17	8.69
Age	44.48	45.21	42.85	42.74	41.20	41.08	42.33	42.56	42.06	42.43	40.80	40.65
Female	0.35	0.35	0.52	0.52	0.45	0.44	0.79	0.80	0.09	0.09	0.84	0.83
Migrants	0.10	0.11	0.13	0.14	0.10	0.10	0.08	0.09	0.09	0.10	0.11	0.12
Education												
Higher(>25)	0.03	0.03	0.08	0.08	0.03	0.03	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.29	0.30	0.50	0.53	0.33	0.32	0.17	0.17	0.10	0.10	0.12	0.12
Intermediate (19-21)	0.59	0.59	0.38	0.37	0.57	0.59	0.72	0.72	0.72	0.73	0.71	0.71
Low (16-18)	0.08	0.08	0.03	0.02	0.05	0.05	0.08	0.08	0.17	0.16	0.13	0.13
None/In Education	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02

	7		8		9	
	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean
Real Hourly Wage	8.85	8.39	10.65	10.24	7.97	7.71
Age	35.81	36.27	44.22	44.21	38.99	38.65
Female	0.67	0.67	0.13	0.13	0.48	0.48
Migrants	0.10	0.10	0.15	0.17	0.16	0.19
Education						
Higher(>25)	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.13	0.13	0.07	0.08	0.08	0.09
Intermediate (19-21)	0.63	0.64	0.68	0.70	0.63	0.63
Low (16-18)	0.10	0.11	0.23	0.21	0.18	0.17
None/In Education	0.13	0.12	0.01	0.00	0.10	0.09

Notes: Entries are for working age(16-64) natives and immigrants for the average real hourl wage, average age, share of female and the share in each education group in 2004 and 2011. Higher educatoin: left full-time education after age 25, High education: left full-time education between age 20-24, Low education: left full-time education between age 16-19, None/Still in education: left education at age 15 or below, or is still in education. Occupation number follows the code assigned in Table B.1. Source: APS 2010-2011

Table B.4. Summary Stats for SOC2000 Occuations in 2010 and 2011

	1		2		3		4		5		6	
	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean
Real Hourly Wage	22.06	21.00	21.21	20.53	16.24	15.70	11.30	11.21	11.70	11.34	9.13	8.70
Age	44.21	44.03	42.66	42.71	41.24	41.44	42.31	42.53	42.04	42.32	40.71	40.55
Female	0.37	0.38	0.47	0.47	0.54	0.54	0.80	0.79	0.09	0.09	0.85	0.84
Migrants	0.10	0.11	0.14	0.14	0.11	0.12	0.08	0.09	0.09	0.10	0.11	0.12
Education												
Higher(>25)	0.03	0.03	0.09	0.09	0.03	0.03	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.31	0.32	0.57	0.57	0.33	0.33	0.17	0.17	0.09	0.10	0.12	0.12
Intermediate (19-21)	0.59	0.58	0.30	0.31	0.58	0.58	0.72	0.72	0.72	0.73	0.71	0.71
Low (16-18)	0.08	0.07	0.02	0.02	0.05	0.05	0.09	0.08	0.17	0.16	0.13	0.13
None/In Education	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02

	7		8		9	
	2010 Mean	2011 Mean	2010 Mean	2011 Mean	2010 Mean	2011 Mean
Real Hourly Wage	8.10	7.97	10.52	10.14	7.98	7.7
Age	35.15	35.92	44.25	44.24	39.06	38.82
Female	0.68	0.68	0.13	0.13	0.47	0.47
Migrants	0.10	0.11	0.15	0.17	0.16	0.19
Education						
Higher(>25)	0.01	0.01	0.01	0.01	0.01	0.01
High (21-24)	0.12	0.12	0.07	0.08	0.08	0.09
Intermediate (19-21)	0.63	0.63	0.69	0.70	0.63	0.63
Low (16-18)	0.11	0.11	0.23	0.21	0.18	0.17
None/In Education	0.14	0.13	0.01	0.00	0.10	0.09

Notes: Entries are for working age(16-64) natives and immigrants for the average real hourl wage, average age, share of female and the share in each education group in 2004 and 2019. Higher educatoin: left full-time education after age 25, High education: left full-time education between age 20-24, Low education: left full-time education between age 16-19, None/Still in education: left education at age 15 or below, or is still in education. Occupation number follows the code assigned in Table B.1. Source: APS 2010-2011

B.3 Average Yearly Migrant Inflows from 2007-2017

Table B.5. Average Yearly Migrant Inflows from 2007-2017 for all specifications

	Average Yearly Percentage Point Change in Migrant Native Ratio 2007-2017					
	All Occ		Low Paid Occ		High Paid Occ	
	Mean	SD	Mean	SD	Mean	SD
	Panel A: Real Hourly Wage SOC 2010 Ordering					
Own Occ	0.93	0.58	1.38	1.13	0.57	0.28
Below Occ	0.99	0.64	1.16	0.80	0.89	0.59
Above Occ	0.80	0.48	1.02	0.81	0.59	0.34
	Panel B: Standard ONS SOC 2010 Ordering					
Own Occ	0.93	0.58	1.38	1.13	0.57	0.28
Below Occ	0.99	0.64	1.55	1.43	0.66	0.25
Above Occ	0.80	0.48	1.15	0.83	0.46	0.25
	Panel C: Real Hourly Wage SOC 2000 Ordering					
Own Occ	0.94	0.58	1.39	1.10	0.58	0.28
Below Occ	1.00	0.64	1.15	0.83	0.91	0.59
Above Occ	0.81	0.51	1.02	0.82	0.59	0.36
	Panel D:Standard ONS SOC 2000 Ordering					
Own Occ	0.94	0.58	1.39	1.10	0.58	0.28
Below Occ	1.00	0.64	1.58	1.38	0.65	0.25
Above Occ	0.81	0.51	1.14	0.88	0.47	0.23
Observations	1287		572		715	

Notes: Entries are for the working age (16-64) average percentage point change in migrant native ratio in occupation-region-time cells from 2007-2017 and it's Standard Deviation (SD), estimated by finding the mean change in migrant native ratio over the period and multiplying it by 100. Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median. Panel A shows the summary statistics when we order SOC 2010 occupations by their hourly wage, Panel B for when we order SOC 2010 occupation by their standard ONS ordering found in the APS, Panel C for when we order SOC 2000 occupations by their real hourly wage, and Panel D for when we order SOC 2000 occupations by their standard ONS ordering. This will change the results when we split it into high and low occupations as the occupation above occupation 6, and the occupation below occupation 5, may be different. SD is standard deviation.

B.4 Mechanisms Real Hourly Wage Ordering

Table B.6. Correlation between education changes and wages - Ordered by Real Hourly Wage SOC 2010

Dependent Variable △ Log Real Hourly Wage	(1) All	(2) High	(3) Low
△ Native Low Educated own occ	-0.314 (0.254)	-0.808 (0.816)	0.0857 (0.254)
△ Migrant Low Educated own occ	0.108 (0.0768)	0.111 (0.0712)	0.0702 (0.154)
△ Native Intermediate Educated own occ	0.0401 (0.212)	0.170 (0.571)	0.0425 (0.229)
△ Migrant Intermediate Educated own occ	-0.0412 (0.0390)	-0.114 (0.0759)	0.0232 (0.0374)
△ High Educated own occ	0.0805 (0.309)	0.247 (0.797)	-0.0930 (0.322)
△ High Educated below occ	0.00439 (0.00801)	0.00998 (0.00916)	-0.00490 (0.0821)
△ High Educated above occ	-0.0116*** (0.00310)	-0.0125*** (0.00455)	0.0246 (0.0576)
Observations	1001	572	429
Control	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Notes: Entries are estimated regression coefficients of the yearly change in native log real hourly wages on the yearly change in the overall proportion with high and higher education in the own, below and above occupations for All, High, and Low SOC 2010 occupations using the real hourly wage ordering for years 2007-2017. Occupations are defined as the 5 highest paid occupations and are below the median average across the 9 occupations and Low Paid occupations are defined as the 4 lowest paid that are below the median and are estimated using 13 government office regions. Additional covariates are controls for migrants and natives separately and include the average age, the proportion with intermediate and low education and year fixed effects. Clustered standard errors are reported in parentheses.
$***p < 0.001$ $**p < 0.01, *p < 0.05$

Table B.7. Correlation between migration and native training - Ordered by Real Hourly Wage SOC 2010

	Training/Education Completed		Training/Education Offered	
	(1) No Control	(2) Control	(3) No Control	(4) Control
△ migration: own occupation	0.0761 (0.0480)	0.0789* (0.0469)	0.0865 (0.0527)	0.0819 (0.0544)
△ migration: below occupation	-0.0335 (0.0220)	-0.0323 (0.0208)	0.00173 (0.0311)	0.00182 (0.0305)
△ migration: above occupation	0.0276 (0.0314)	0.0280 (0.0334)	-0.00913 (0.0473)	-0.00781 (0.0468)
Observations	1001	1001	1001	1001
Time FE	Yes	Yes	Yes	Yes

Notes: Entries are estimated regression coefficients of the yearly change in the proporiton of natives who have taken or have been offered but rejected job related training or education on the yearly change in the employed migrant to native ratio, $\triangle m_{ort}$ in the own, below and above SOC 2010 occupations using the real hourly wage ordering for years 2007-2017. All estimations include 9 occupation groups ordered by real hourly wages and are estmated using 13 government office regions. Additional covariates are controls for migrants and natives separately and include the average age, the proportion with higher, high, intermediate and low education and year fixed effects. Clustered standard errors are reported in parentheses. $***p < 0.001$ $**p < 0.01$, $*p < 0.05$