Universitat Pompeu Fabra
Departament d'Economia i Empresa

DOCTORAT EN ECONOMIA, FINANCES I EMPRESA

Essays in the Economics of Migration

Tesi Doctoral de:
Sébastien Willis

Directors de la tesi:
Albrecht Glitz
Ruben Enikolopov

TESI DOCTORAL UPF / 2020

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Albrecht and Ruben, without whose support, encouragement, advice, and occasional prodding, this thesis would certainly never have existed. I would also like to thank the members of the UPF LPD group and the CREI lunchtime seminar for their feedback and suggestions along the way, and Marta Araque and Laura Agustí for their support and patience throughout.

Merci vielmal Flavio for the years of moral support and for taking my ramblings about fixed effects seriously. *Gracies* also to Ilja, Ana and Ana (in no particular order), Dani, Adilzhan, Jared, Cate, Niklas, Niko and all my UPF friends and colleagues who've helped and supported me in all kinds of tiny ways ; no one gets through this thing alone.

Thank you Melike, for your smile throughout, for making everything fun, and for your patience at the end.

And to my parents, *merci tout court*.

Sébastien Willis
Barcelona
December 2020

Abstract

This thesis consists of three independent articles. In the first chapter, I test whether tipping points can explain observed workplace segregation between immigrants and natives in Germany over the period 1990-2010. I reject the hypothesis of tipping dynamics. Furthermore, I show that traditional tests of tipping points based on Regression Discontinuity Designs tend to over-reject the null hypothesis of no tipping relative to a procedure that correctly accounts for uncertainty in the location of the tipping point.

In the second chapter, I study the effect of workplace segregation on the outcomes of immigrants in Germany. Starting one's career in a workplace employing relatively many conationals lowers an immigrant's subsequent employment rates, but does not affect wages conditional on employment. The effect appears to be driven by lower-quality coworker networks rather than differential accumulation of human capital as a result of the initial place of work.

In the third chapter, I study the effect of rural-urban migration on labour market outcomes in Indonesia. I find that urban migrants are more likely to be employed than both siblings who stayed behind and observationally similar urban natives in the short-run and experience more rapid occupational upgrading than their siblings in the long run.

Resum

Aquesta tesi consta de tres articles independents. Al primer capítol, comprovo si els punts d'inflexió poden explicar la segregació entre immigrants i nadius observada entre diferents llocs de treball a Alemanya durant el període 1990-2010. Rebutjo la hipòtesi de la existència de punts d'inflexió. A més, mostro que els tests tradicionals de la existència de punts d'inflexió, basats en dissenys de regressió discontínues, tendeixen a rebutjar excessivament la hipòtesi nul·la respecte a un contrast que té en compte la incertesa en la localització del punt d'inflexió.

Al segon capítol, miro l'efecte de la segregació en el lloc de treball per als immigrants a Alemanya. Començar la carrera professional en un lloc de treball on hi ha relativament moltes persones del mateix país d'origen redueix les taxes d'ocupació posteriors d'un immigrant, però no afecta els salaris en el cas de treballar. L'efecte sembla estar explicat per xarxes de companys de treball de menor qualitat, i no per una acumulació diferencial de capital humà com a resultat del lloc de treball inicial.

El tercer capítol s'enfoca en l'efecte de la migració rural-urbana en les oportunitats laborals dels emigrants en el context d'Indonèsia. A curt termini, els emigrants es beneficien de taxes d'ocupació més elevades comparat amb els seus germans que no van migrar i amb persones semblants que viuen a la ciutat, però no hi ha cap efecte a llarg termini. Els emigrants també es beneficien d'una millora més permanent en l'ocupació respecte als seus germans, amb una situació semblant a la dels residents urbans comparables.

Preface

This thesis consists of three independent articles on the economics of migration. The three articles share an empirical approach to studying the labour market outcomes of immigrants, whether they be international migrants in developed countries, or internal migrants in developing countries.

Segregation of immigrants across workplaces has been widely documented, however the causes and consequences of segregation remain subject to conjecture. In the first chapter, I use social security data for the period 1990-2010 to study whether tipping points in the composition of industries can explain observed patterns of segregation across industries by ethnicity in Germany. I consider two tests for the existence of tipping points in the composition of local industries' workforces, one based on a regression discontinuity design (RDD) around a candidate tipping point, the other based on a threshold regression that includes an unknown breakpoint. I find only limited support for the existence of tipping dynamics in native employment flows using RDD methods and no evidence when estimating a threshold regression. The RDD evidence is strongest for the period 1990-1995, when immigrant inflows to Germany were largest. Furthermore, my findings suggest that inference methods previously used to test for the existence of tipping points in labour markets may have a tendency to over-reject the null of no tipping points. Taken together, my results may be cause for some scepticism about the existence of tipping points in labour markets.

In the second chapter, I turn to the consequences of workplace segregation. I use survey data matched to administrative records to study the effect of segregation in an immigrant's first job on her subsequent labour market outcomes. I argue that controlling for the wealth of pre-migration characteristics recorded in my survey data, not typically available in studies of immigrant outcomes, is sufficient to account for selection into high-conational firms. Both OLS and semi-parametric estimates indicate that a one-percentage-point increase in the share of conationals in an immigrant's first job is associated with 0.16–0.18-percentage-point lower employment rates in the medium- to longer-term, while there is no clear evidence of an earnings effect. Formal tests show that the results are robust to selection on unobservables. Differences in human capital acquisition do not appear to explain the employment effect, while there is some evidence that it is explained by differences in the quality of social network induced by differences in the initial workplace.

In the third and final chapter, I study rural-urban migration in the developing world. There are large earnings gaps between urban and rural locations in the developing world, raising the question of why more rural residents do not emigrate. I use an event study comparing urban migrants with siblings who stayed back to estimate the effect of urban migration on the labour market outcomes of urban migrants in Indonesia. I find that migrants experience a positive but transient employment boost and more permanent occupational upgrading relative to their siblings. I then compare migrants to comparable urban natives, identified using a matching procedure, to establish that, unlike international migrants, urban migrants do not experience a labour market penalty relative to urban natives. The evidence suggests that barriers to rural-urban migration are higher than one would infer from simple comparisons of urban and rural wages.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

TIPPING POINTS AND THE DYNAMICS OF ETHNIC SEGREGATION ACROSS INDUSTRIES IN GERMANY

1.1 INTRODUCTION

Recent research has established that immigrants and natives are highly segregated across industries and workplaces in many developed economies (Hellerstein and Neumark, 2008; Åslund and Skans, 2010; Andersson et al., 2014; Glitz, 2014). In the case of Germany, where the foreign-born made up 12.8 per cent of the population in 2008 (OECD, 2020), 40 per cent of immigrants would have needed to change firms to achieve a degree of segregation consistent with a random assignment of workers to firms. Even accounting for differences in location, education, and gender between immigrants and natives, 26 per cent of immigrants would have needed to change firms (Glitz, 2014).

Workplace segregation unexplained by observed characteristics suggests factors of production are misallocated, which can have large negative consequences for aggregate productivity and output (Hsieh et al., 2019).

At the individual level, segregation across workplaces or industries could help explain the widely-studied persistence of employment and wage gaps between immigrants and natives (e.g. Lubotsky, 2007; Sarvimäki, 2011) and the fact that immigrants tend to work at lower-paying firms (Aydemir and Skuterud, 2008; Barth et al., 2012). Since coworker networks are an important source of information and referrals in the labour market (Cingano and Rosolia, 2012; Eliason et al., 2019; Glitz and Vejlin, 2020), segregation across workplaces or industries could restrict immigrants' access to better-paying jobs and firms if it means they lack the native coworkers necessary to land these jobs.

The causes of observed segregation, however, are not yet fully understood. Three broad theoretical explanations have been advanced. First, the oldest set of explanations show how discrimination on the part of employers towards certain types of workers, such as immigrants, leads to workplace segregation. Such discrimination could be for reasons of taste (Becker, 1957), or for statistical reasons (Aigner and Cain, 1977). Second, there is a long tradition arguing that spillovers in either consumption, in the tradition of social interaction models (Schelling, 1971)—immigrants might prefer to work with other immigrants and natives with other natives—or in productivity, say due to communication costs (Lazear, 1999a), could optimally lead to the formation of homogeneous workplaces. Finally, the most recent group of explanations focuses on how segregated social networks—the tendency, for example, of immigrants to befriend other immigrants—can lead to segregation in labour market outcomes, including the place of work (see Jackson et al., 2017, for a review).

The objective of this chapter is to empirically examine the second of these proposed explanations, that is, to search for evidence that spillovers, particularly spillovers in preferences, can lead to workplace segregation. I do so in the context of Germany in the period 1990-2010. Preference spillovers arise when the composition of the workplace or industry directly affects individuals' utility. This creates a strategic interaction between individuals as they choose which industry to supply their labour to. The preference spillovers I study here could be caused by a simple distaste for working with immigrants, or they could be caused by career concerns. Such concerns could arise if immigrants are a worse source of information about the labour market and referrals, leading to lower job mobility for their coworkers, or if an increasing immigrant share is taken as a signal that an industry has experienced a negative productivity shock (c.f. Goldin, 2014). My results do not depend on the specific cause of preference spillovers.

There are several reasons to focus on preference spillover-based explanations of segregation. First, preference spillovers are one of the leading explanations of observed patterns of residential segregation (Schelling, 1971; Cutler et al., 1999; Becker and Murphy, 2000; Card et al., 2008). Second, while systematic evidence over time is relatively scarce, in 2017 only 37 per cent of Germans stated they would be "totally comfortable" having an immigrant as a work colleague, similar to the proportion (36 per cent) stating that they would be totally comfortable having an immigrant as a neighbour (European Commission, 2018).[1] These two facts suggest that the preference spillovers that help explain residential segregation might also be a leading explanation for workplace segregation. Third, a common feature of many models of preference spillovers in residential choice is that they generate stark predictions about the dynamics of neighbourhood composition, which readily lend themselves to empirical testing. Specifically, neighbourhood composition is often predicted to follow a "tipping" dynamic; once the immigrant share exceeds a certain threshold it is predicted to rapidly increase towards one. This stark prediction makes it relatively straightforward to test whether preference spillovers might be a cause of workplace segregation.

I proceed by presenting a simple model of local industry workforce composition that adapts the model of neighbourhood composition of

[1] The other options were "somewhat comfortable", "somewhat uncomfortable", "totally uncomfortable", or "don't know". Across the EU, the share "totally comfortable" was 43 per cent for neighbours and 44 per cent for colleagues.

Card et al. (2008). In this model, the labour supply to an industry in a local labour market depends on both the wage offered in the industry and the composition of the industry's workforce, a form of social preference in the spirit of Schelling (1971). An equilibrium is characterised by the immigrant share in the local industry. The model has multiple equilibria for a given relative supply of immigrants, though typically only one stable integrated equilibrium, where the immigrant share is greater than zero and less than one. Importantly, there is a discontinuity in the response of the equilibrium immigrant share to changes in the relative supply of immigrants. Following Card et al. (2008), a tipping point is defined as the maximum of the set of stable integrated equilibria.

This definition of a tipping point implies a discontinuity in the evolution of the immigrant share over time. Local industries below the tipping point, in the interior of the set of stable integrated equilibria, should see their immigrant share change little in response to small changes in the relative supply of immigrants. However, once the immigrant share in a local industry is above the tipping point, the immigrant share should start increasing rapidly as the industry shifts towards the segregated equilibrium. This definition of a tipping point differs from the traditional definition of a tipping point as a single unstable equilibrium (Schelling, 1971; Becker and Murphy, 2000).

I carry out two distinct empirical tests for the presence of such tipping dynamics in workforce composition. In the first test, I implement the two-step procedure proposed by Card et al. (2008). In the first step, I identify candidate tipping points in the composition of local industries in West Germany for the periods 1990-1995, 1995-2000, 2000-2005, and 2005-2010, using an ad hoc search procedure proposed by Card et al. (2008). I allow the location of the candidate tipping point to vary across local labour markets. In the second step, I apply regression discontinuity design (RDD) techniques (Imbens and Lemieux, 2008; Lee and Lemieux, 2010) and look for evidence of discontinuities at the identified candidate tipping points. The evidence from this approach is mixed. There is some evidence of tipping, in the form of either native flight from or native avoidance of relatively high immigrant-share industries, during the first two periods, when net immigration to Germany was high, though no evidence in the later two periods, when net immigration was low and Germany experienced several recessions.

Since this evidence is not conclusive, I conduct a second test, applying formal techniques for identifying and testing for the existence of breakpoints in a conditional expectation function using a threshold regression (Hansen, 1996, 2000). Threshold regressions have not yet to my knowledge been formally used to test for the existence of tipping points. This test leads me to reject the hypothesis of a discontinuous change in the evolution of workforce composition in all periods. I also check for evidence of tipping points in the composition of individual establishments, rather than industries, using a threshold regression. Here too I fail to find evidence of tipping dynamics that could explain observed workplace segregation.

More careful comparison of RDD and threshold regression approaches suggests that the discrepancy between the two is due to a tendency of RDD-based tests for the existence of tipping points using bootstrap standard errors to over-reject the null of no discontinuity. Specifically, standard errors for the size of the discontinuity in the RDD approach are typically calculated using the bootstrap. However, the location of the tipping point is treated as fixed across bootstrap samples, reducing the variability of the discontinuity relative to a procedure where both the location of the tipping point and the size of the discontinuity at the tipping point can vary over bootstrap samples. Since the location of the true tipping point is unknown, the latter procedure is the correct one for conducting inference on the size of any discontinuity in the outcome at the tipping point.

The use of bootstrap standard errors was originally suggested by Card et al. (2008), although they adopt a different inference procedure in their main results, based on splitting the sample into separate subsamples for identifying the location of the tipping point and testing for a discontinuity. The sample splitting approach requires relatively abundant data, as a result of which more recent work on tipping points in labour markets (Pan, 2015) has relied on bootstrapped standard errors when testing for tipping points. My results suggest some scepticism may be appropriate when evaluating the results of these bootstrap-based tests for the existence of tipping points in labour markets.

The chapter is structured as follows. In the following section I review the literature on segregation and tipping points in firms and neighbourhoods, where the question has been more extensively studied. In Section 1.3 I outline a model of workplace segregation and show how discontinuities in industry workforce growth arise. In Section 1.4 I outline the empirical implications of the model and the two tests I propose for the existence of tipping points. In Section 1.5 I present the results of the two tests. Finally, Section 1.6 concludes.

1.2 Literature

The best-known studies on segregation have tended to focus on residential segregation. Extensive residential segregation has been documented by ethnicity in the US (Cutler et al., 2008), and Europe (Semyonov and Glikman, 2009). Early papers on the consequences of residential segregation found that it was negatively associated with the outcomes of minorities (Cutler and Glaeser, 1997) and immigrants (Borjas, 1995), while nevertheless recognising that segregation arises endogenously and might still be optimal from the individuals' perspective (Borjas, 1998). More recent empirical evidence has emphasised the role of non-random selection in driving observed negative findings (Edin et al., 2003; Damm, 2009), noting that segregation can improve job-finding probabilities and consequently employment and wages in the short run, though residential segregation appears to lower immigrants' human capital acquisition in the the long run (Battisti et al., 2018).

The literature on workplace segregation developed in parallel to the literature on residential segregation, though it is smaller in comparison. Earlier papers suffered from limited access to disaggregated data and were constrained to show evidence of segregation by occupation or industry (Albelda, 1986) or for historical periods (Higgs, 1977). The proliferation of large-scale firm datasets drawn from administrative records in the last two decades has now allowed researchers to document significant workplace segregation in the US (Hellerstein and Neumark, 2008; Andersson et al., 2014), Sweden (Åslund and Skans, 2010), and Germany (Glitz, 2014). Glitz (2014) shows that workplace segregation also correlates with immigrants' economic outcomes: immigrant cohorts become less segregated from natives with time spent in Germany, just as their wages converge to those of natives.

There is a long tradition of models that seek to explain observed patterns of segregation by appealing to social interactions models, (Schelling, 1971, 1978; Becker and Murphy, 2000). In these models, a small preference for majority-dominant units (neighbourhoods, firms, schools, etc.) can lead to extensive segregation across units. These models are characterised by a multiplicity of equilibria, some of which may be unstable, potentially leading the observed composition of integrated units to shift rapidly to a stable, segregated equilibrium when subject to some shock. In contrast, Card et al. (2008, 2011) propose a model where integrated equilibria are stable, however only low immigrant shares can be supported in equilibrium; if the relative demand of immigrants for housing in a neighbourhood increases too much, no integrated equilibrium will exist.

While Easterly (2009) claimed to find no evidence of tipping in the composition of US neighbourhoods, there is growing evidence of the empirical relevance of tipping points. Card et al. (2008) develop a reduced-form method, applied in this chapter, where a tipping point is understood as an immigrant share at which there is a discontinuity in the expected change in the share of natives in a neighbourhood, and find ample evidence of tipping in the composition of US neighbourhoods.

Aldén et al. (2015) apply the same method and find similarly clear evidence of tipping points in the composition of Swedish neighbourhoods. Böhlmark and Willén (2020) use neighbourhood tipping points identified by a Card-style procedure to study the effect neighbourhood composition on children's educational and future labour market outcomes in Swedish metropolitan areas, arguing that the identified tipping points can be treated as an RDD-style cutoff in the immigrant share of the neighbourhood where a child grows up when studying these later outcomes. In the labour market, the method of Card et al. has been used to identify tipping points in occupational composition by gender in the US (Pan, 2015), which is the paper most similar to the present chapter. However, the method has not yet been used to study ethnic segregation in the labour market, the focus of this chapter. Caetano and Maheshri (2017) have also developed a more structural method for identifying tipping points, and use it to identify tipping points in the composition of schools' student bodies in Los Angeles.

1.3 THEORY

1.3.1 *Unit of analysis*

The model of tipping I propose here builds on the work of Schelling (1971) and Card et al. (2008, 2011), where members of the majority group experience increasing disutility as the share of minority individuals in their unit of analysis increases. In the papers just cited, the unit of analysis is the neighbourhood. Before showing how such a model of social preferences can lead to tipping dynamics and explain patterns of workplace segregation, I therefore need to establish the relevant unit of analysis in which natives might experience disutility from working with a larger share of immigrants.

In this chapter I focus on tipping in the composition of 3-digit industries within local labour markets, a unit of observation I refer to as a *local industry*, or simply *industry*. There are theoretical, empirical, and practical reasons to focus on local industries.

Theoretically, the choice of unit of analysis will depend on the reasons for which natives experience disutility from working with immigrants. If the disutility were experienced in personal interactions with colleagues, the production unit (team, small plant) would be the natural unit of analysis. If, however, the disutility arises because increasing the immigrant share lowers the prestige of a sector, then the local industry will be a more natural choice. Goldin (2014) has argued that an increasing female share can be taken as a signal of a negative productivity shock in an occupation, lowering its prestige. The same is plausibly true of the immigrant share in an industry. Certainly there is historical evidence that the low prestige associated with jobs in certain industries led natives to avoid employment in those industries with the resulting labour shortage made up for through immigration (Noiriel, 1988). Finally, if tipping dynamics are driven by the decisions of new workers searching for jobs, the fact that the immigrant share is more readily observable to an outsider at the level of a local industry than at a firm also suggests that local industries are a more natural level at which to build a test for evidence of tipping dynamics.

Empirically, segregation is observed across teams, plants, firms, local industries, national industries, or occupations, and social preferences could potentially be present and cause tipping dynamics in the composition of any of these units. However, segregation across local industries does account for a large proportion of total workplace segregation. Table 1.1 reports the index of coworker segregation, defined by Hellerstein and Neumark (2008) as the excess probability that an immigrant has of working with other immigrants, relative to a native, for West Germany in 1990-2010. This index can be normalised to account for differences in the distribution of immigrants and natives across larger units of aggregation, such as regions, yielding what is known as an *effective* index of coworker segregation. Conditioning the index on the distribution of workers over local labour markets and 3-digit industries explains around 45 per cent of observed segregation. Indeed, segregation across 3-digit industries accounts for around 40 per cent of observed segregation *within* local labour markets. Explaining segregation across industries within local labour markets would therefore go a long way towards explaining observed patterns of total workplace segregation.

Table 1.1: Index of coworker segregation

	1990 ICS	1995 ICS	2000 ICS	2005 ICS
Unconditional	0.15	0.17	0.18	0.17
Conditional on industry	0.12	0.13	0.14	0.13
Conditional on location	0.14	0.16	0.17	0.16
Conditional on location and industry	0.08	0.09	0.10	0.10
Establishments	523246	548910	723572	747317

Note: Indexes of coworker segregation calculated from the *Betriebshistorikpanel*, including all establishments in West Germany hiring two or more employees. The conditional indexes condition on either three-digit industry (NACE Rev. 1), local labour market, or both.

Finally, as a practical consideration, establishments tend to be relatively small, with over 70 per cent of firms in my dataset employing fewer than 10 employees, implying that both the size and composition of workforces can vary widely over time. This introduces considerable noise to the data, potentially making it harder to empirically detect tipping dynamics. Furthermore, there is selection out of the sample over time as establishments close, potentially introducing bias since tipping also takes place over time. Focusing on local industries will allow me to overcome these practical difficulties. In a robustness check I will nevertheless confirm that firm-level estimates are consistent with my local industry-level estimates.[2]

[2]Note that there is no straightforward logical relationship between tipping at the industry level and at the firm level. Industry-level tipping does not imply firm-level tipping, since it could occur through the entry of high immigrant-share or the exit of

1.3.2 *A model of tipping*

In this section I adapt the model of Card et al. (2008, 2011) of neighbourhood composition in the presence of social interactions to study the composition of local industries' workforces. The model is static and partial equilibrium. A representative, nondiscriminating firm hires two types of workers, immigrants and natives, denoted $j \in \{I, N\}$, which it treats as perfectly substitutable in production. The industry's size, and hence the representative firm's size, is taken as given, so the total workforce is normalised to equal one. The inverse supply of type j is given by $\omega^j(n_j, s)$, a primitive of the model. Crucially, the inverse supply depends not only on the quantity of workers of type j hired, n_j, but also on the share of immigrants in the firm, s.

The partial derivatives $\partial\omega^j(n_j, s)/\partial n_j$ are assumed to be weakly positive, that is, for a constant immigrant share, the firm needs to raise wages to hire more workers of a given type. The partial derivative $\partial\omega^j(n_j, s)/\partial s$ represents the social interaction effects. In particular, I assume that $\partial\omega^N(n_N, s)/\partial s > 0$ for s greater than some threshold; that is, as the immigrant share in the firm increases beyond some threshold, the firm needs to pay a higher wage to hire a given quantity of natives. Under the normalisation that the total workforce is one, we have $n_N = 1 - s$, and the derivative of the native inverse supply function with respect to the migrant share will be

$$\frac{d\omega^N}{ds} = -\frac{\partial\omega^N}{\partial n_N} + \frac{\partial\omega^N}{\partial s}. \tag{1.1}$$

Under the previous assumptions, the first term will be negative above some threshold, while the second term will be positive. I follow Card et al. (2008) in assuming that the social interaction effect is sufficiently strong such that $d\omega^N/ds > 0$ for high levels of s, i.e. supply of natives

high native-share firms as the industry passes the tipping point. Similarly, tipping at the firm level need not imply tipping at the industry level if individual firm tipping only implies a reallocation of a fixed pool of workers within the industry.

$n_N = 1 - s$ is downward sloping for low levels of n_N and only becomes upward-sloping as n_N rises and the immigrant share s falls below a certain threshold. I also assume for simplicity that $d\omega^I/ds > 0$ for all $s \in (0, 1)$, that is that supply of immigrants is upward-sloping for all values of n_I.[3]

There are multiple ways one could interpret the social interaction effects captured by the assumption that $\partial\omega^N(n_N, s)/\partial s > 0$. The simplest way, consistent with the original model of Card et al. (2008) and the tradition of social interactions models going back to Schelling (1971), is to interpret this as a consumption externality. Natives experience disutility from working with immigrants, so the marginal native worker will become unwilling to work at the firm if the immigrant share increases.

The source of this disutility could be a simple distaste or discomfort experienced by individual natives when working with immigrants. Alternatively, the disutility could arise from dynamic considerations, if natives believe that working with immigrants will harm their future job-finding prospects and earnings, say, because immigrants are not a good source of referrals or information about job openings. The disutility might also arise indirectly, if an inflow of immigrants into an industry provides a negative signal about productivity in the industry, as in the pollution model of Goldin (2014), lowering the prestige of working in the industry.

Alternatively, one could also interpret the social interaction effect as a productivity externality, reinterpreting n_N as the effective supply of natives. Under this interpretation, an increase in the immigrant share lowers the productivity of natives; to keep a constant effective

[3]There is therefore an asymmetry in the strength of the social interaction effects between immigrants and natives that drives an asymmetry in the shape of the inverse supply curves of migrants and natives. This asymmetry is also present in the model of Card et al. (2008). The empirical predictions of the model can still be derived when social interactions cause immigrant inverse supply to be downward sloping for low values of s; what is strictly necessary however is that the inverse supply curve of immigrants be flatter than the inverse supply curve of natives, i.e. $d^2\omega^I/ds^2 < d^2\omega^N/ds^2$, for all $s \in (0, 1)$.

supply of native workers, the firm must raise the wage offered to hire more natives. This interpretation is consistent with recent evidence on negative productivity spillovers between immigrants and natives in certain firms (Glover et al., 2017), however it would also complicate the derivation of Equation (1.1), since now $n_I \neq s$, so I do not entertain it further here.

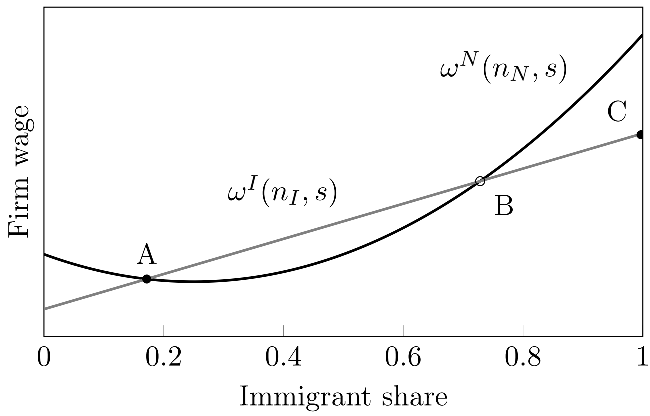
The inverse supply curves of immigrants and natives are plotted in Figure 1.1. As $s = n_I = 1 - n_N$, the supply of immigrants increases moving to the right on the x-axis, while the supply of natives increases moving to the left on the x-axis. At an integrated equilibrium, where both types of workers are employed in the industry, the wages paid to both types of workers must be equal, since the firm is assumed to be non-discriminating. Again under the normalisation that the total workforce is one, equilibrium therefore requires

$$\omega^N(1-s,s) = \omega^I(s,s). \tag{1.2}$$

As the inverse supply curves are drawn, there are three equilibria. Equilibrium A is stable in the sense that a small increase in the firm's minority share raises the wage that must be paid to immigrants above the wage paid to natives, so the firm hires natives until it returns to the equilibrium at A. The same remark holds *mutatis mutandis* for a decrease in the minority share at A or at C. Equilibrium B is, however, unstable. After a small increase in the immigrant share from B, the wage demanded by natives is greater than the wage demanded by immigrants, the firm will replace natives with immigrants until it reaches the equilibrium at C.

In Figure 1.2 I plot what happens as the supply of immigrant workers to the firm increases, say, as a result of an inflow of immigrants to the local labour market where the firm is located. Suppose the firm is initially in equilibrium at E_1. As the supply of immigrants increases, their inverse supply curve shifts downward, and the equilibrium gradually moves to the right. However if the inflow of immigrants continues, the point of tangency E_2 will eventually be reached, which is stable with

Figure 1.1: Immigrant and native inverse labour supply

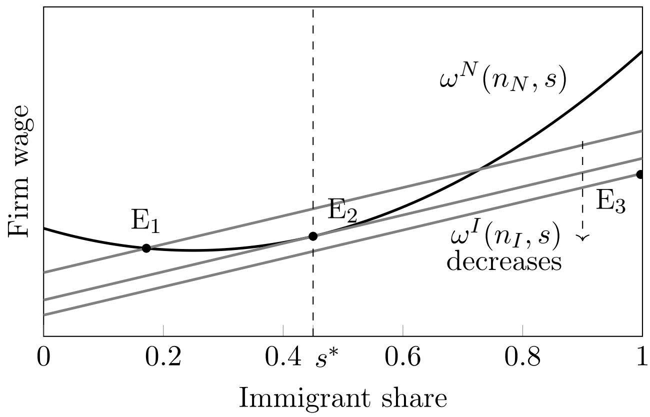


Notes: Immigrant and Native inverse labour supply to the firm with three equilibria. A and C are stable, B is unstable.

respect to decreases in the immigrant share, but unstable with respect to increases. If there are any further increases in the supply of immigrant workers, no integrated equilibrium will exist, the only equilibrium will involve the firm hiring only immigrants, as at point E_3. Traditional social interaction models such as Becker and Murphy (2000) identify the unstable equilibrium B in Figure 1.1 as a tipping point. Here, however, I follow Card et al. (2008) in defining the tipping point as the maximum possible immigrant share in an integrated equilibrium. In Figure 1.2, this is the immigrant share s^*, associated with the equilibrium E_2.

Two caveats are worth noting with this model. First, it does not account for the distribution of immigrants across industries, only the composition of a single industry. Implicitly I assume that the natives who leave the industry after the tipping point is exceeded would either prefer to be unemployed than keep working in a high-immigrant-share industry, or are able to find jobs in other industries that have not faced a similar supply shock. The latter scenario could arise if the size of other industries is not similarly constrained, or if the immigrant supply

Figure 1.2: Effect of increasing supply of immigrant labour



Notes: Increasing supply of immigrant workers shifts their relative supply outwards, decreasing the wage demanded for any value of s. The equilibrium immigrant share starts at E_1 and shifts right as the inverse supply of migrants increases. The equilibrium E_2 is the maximum integrated equilibrium, the associated migrant share is s^*. If the supply of immigrant workers increases further, the firm will jump to the segregated equilibrium E_3, hiring only immigrants.

shock is specific to a single industry, perhaps because it is mostly located in the neighbourhood where newly arrived immigrants settle. Second, social interaction models are typically thought to lead to an inefficiently high degree of segregation across neighbourhoods, because agents cannot coordinate on where to locate. The model presented here, by only considering a single representative firm, is silent about the potential welfare consequences of such social preferences. It has traditionally been argued that firms, by internalising any spillovers across workers arising from their hiring decisions, choose a socially optimal degree of segregation (Becker and Murphy, 2000). However, these arguments do not account for the possibility that workplace segregation could be dynamically inefficient, if it keeps immigrants from developing the network necessary to move up the job ladder.

1.3.3 *Dynamic implications*

While the model presented in the previous section is static, it is still possible to use it to make dynamic predictions about the composition of the representative firm's workforce.

Consider a firm whose initial static equilibrium immigrant share is $\bar{s}_0 < s^*$, where s^* is the tipping point defined previously as the immigrant share associated with the maximum possible integrated equilibrium. The firm experiences a small increase in the supply of immigrants, i.e. a fall in the wage immigrants need to be paid, $\Delta\omega^I(n_I, s) < 0$, between period 0 and period 1.[4] There will be some $r \in (0, s^*)$ such that if $\bar{s}_0 \in [0, s^* - r)$, the firm's new equilibrium will be at $\bar{s}_1 \in (0, s^*]$, whereas if $\bar{s}_0 \in [s^* - r, s^*]$, the increase in the immigrant supply takes the firm beyond the point of tangency at E_2 in Figure 1.2 and the new equilibrium will be $\bar{s}_1 = 1$. As the increase in the immigrant supply $\Delta\omega^I(n_I, s)$ becomes infinitesimally small, r also approaches zero. Note that no firm can initially be at an equilibrium at $\bar{s}_0 \in (s^*, 1]$ except for at $\bar{s}_0 = 1$, where a small increase in the supply of immigrants will have no effect on the equilibrium.

Assume that the firm myopically adjusts its immigrant share in response to changes in the supply of immigrants such that the immigrant share s_t remains close to its equilibrium value. To allow for the possibility that search or other labour market frictions prevent the immigrant share from fully adjusting within a single period to a new equilibrium value as the supply of immigrants changes, I use the notation s_t to refer to the observed immigrant share at a point in time, to distinguish it from the static equilibrium at that point in time, $\bar{s}_t$. For an observed $s_0 \in [0, s^* - r)$, the observed increase in the immigrant share Δs_1 in response to the increase in the immigrant supply $\Delta\omega^I(n_I, s)$ will be small. However, for $s_0 \in [s^* - r, s^*]$, $\Delta\omega^I(n_I, s)$ will cause a large observed Δs_1,

[4] The discussion here in fact holds for an increase in the relative supply of immigrant, $\omega^N(n_N, s) - \omega^I(n_I, s)$. However, to simplify the discussion I assume the supply of natives is fixed and only the supply of immigrants varies.

as the firm converges to the new equilibrium at $\bar{s}_1 = 1$. For firms initially at $s_0 \in (s^*, 1)$, the tipping process is already underway, and one should expect to see $\Delta s_1 > 0$ and larger the closer the firm is to s^*. There will therefore be a discontinuity in Δs_1 around the tipping point s^*. We will observe Δs_1 to be small and positive for s_0 to the left of the tipping point and large and positive for s_0 close to or beyond the tipping point.

Whilst the foregoing discussion restricts attention to the case of an increase in the immigrant supply, where the discontinuity appears clearly, the discontinuity will also exist in the case where there is a decrease in the immigrant supply. This is because once a firm has started tipping and $s_0 \in (s^*, 1]$, a small decrease in the supply of immigrants will typically not reverse the tipping process, implying that for these firms too $\Delta s_1 > 0$. The condition for tipping to continue after a decrease in the immigrant supply is for the marginal immigrant to continue to accept a lower wage than the marginal native, which is more likely to be satisfied the smaller the decrease in the immigrant supply or the further to the right of s^* the firm initially finds itself. On the other hand, for a firm that is close to tipping, but where $s_0 < s^*$, a small decrease in the immigrant supply will lead to a small decrease in the immigrant share in the firm.

Combining these observations about the effect of increases and decreases in the immigrant supply on the firm's immigrant share, one can conclude that there will be a discontinuity in the expected change in the immigrant share as a function of the base-year immigrant share:

$$\mathrm{E}[\Delta s_t | s_{t-1}] = \mathbf{1}(s_{t-1} < s^*) g(s_{t-1}) + \mathbf{1}(s_{t-1} \geq s^*) h(s_{t-1}) \tag{1.3}$$

where $\lim_{\varepsilon \to 0^+} h(s^* + \varepsilon) - g(s^* - \varepsilon) > 0$. $h(s_{t-1}) > 0$, while the sign of $g(s_{t-1})$ will depend on whether firms more commonly face increases or decreases in the immigrant supply. The existence of a discontinuity in $\mathrm{E}[\Delta s_t | s_{t-1}]$ at the tipping point s^*, which does not depend on whether the immigrant supply is increasing or decreasing, is the key dynamic implication of the model I will test in the empirical analysis below.

The location of the tipping point s^* will depend on the value of the partial derivatives of the inverse supply functions and, in particular, the strength of native distaste for immigrants, measured by the partial derivative $\partial\omega^j(n_j, s)/\partial s$. If the value of the partial derivatives of the inverse supply functions is the same across labour markets, then the tipping point will also be the same for different labour markets. Both Card et al. (2008) and Aldén et al. (2015) assume different tipping points for different residential markets, while Pan (2015) assumes the location of tipping points in labour markets varies regionally. The strength of native distaste of immigrants, which determines the location of the tipping point, likely varies with the level of historical exposure to immigrants, which varies across locations. I will therefore follow the previous literature and assume the location of the tipping point varies across local labour markets in the two-step estimates presented below. However, I will also consider specifications where the tipping point is assumed to be common to all labour markets, which does not alter my conclusions.

While the model presented above assumes that the size of the industry is fixed, in reality industries vary in size, and grow and contract over time. Let the number of immigrants employed in an industry at time t be I_t, the number of natives be N_t, and the total number of employees be L_t. The immigrant share is then defined as $s_t = I_t/(N_t + I_t)$. Rather than focussing on changes in s_t, the main dependent variable in my empirical specifications is the five-year growth in the industry's native and immigrant workforces, normalised by the total workforce in the base year, $\Delta n_{t,t+5} = (N_{t+5} - N_t)/L_t$ and $\Delta i_{t,t+5} = (I_{t+5} - I_t)/L_t$. This has the advantage relative to using $\Delta s_{t,t+5} = I_{t+5}/L_{t+5} - I_t/L_t$ as a dependent variable of keeping the denominator fixed, focusing on changes in workforce composition not driven simply by changes in workforce size.

An analogous version of Equation (1.3) asserts that the growth of the native workforce, $\Delta n_{t,t+5}$ and the growth of the immigrant workforce, $\Delta i_{t,t+5}$ are smooth functions of the base-year migrant share, s_t, except at the tipping point s^*. Here, $\Delta n_{t,t+5}$ will fall discontinuously, and $\Delta i_{t,t+5}$

will increase discontinuously. A discontinuous fall in $\Delta n_{t,t+5}$ is consistent with either native flight from, or native avoidance of industries past the tipping point.

1.4 EMPIRICAL APPROACH AND DATA USED

Any method testing for the existence of a tipping point in industry workforce composition needs to reckon with the fact that the theoretical tipping point s^* is unknown. I consider two different approaches to doing this: the original method of Card et al. (2008) and a method based on a threshold regression (Hansen, 1996, 2000).

1.4.1 *Two-step empirical specification*

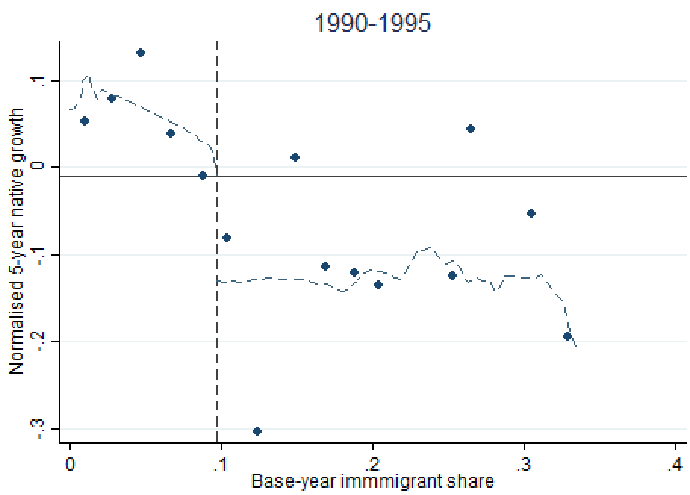
Card et al. (2008) propose a two-step procedure to test for the existence of a tipping point. In the first step, they propose an ad hoc method, based on the literature on structural breaks, to identify a candidate tipping point from the data. The method works by approximating the change in the dependent variable, in this case normalised native workforce growth in industry j, $\Delta n_{j,t,t+5}$, as a constant function with a single discontinuity at some unknown break point,

$$\Delta n_{j,t,t+5} = a_{c(j),t} + \delta_{c(j),t}\mathbf{1}(s_{j,t} \geq s^*_{c(j)t}) + \varepsilon_{jt} \tag{1.4}$$

The tipping point $s^*_{c(j)t}$, which is assumed to be specific to the local labour market $c(j)$ in which industry j operates and base year t, is chosen as the value in $[0, 60]$ that maximises the R^2 of Equation (1.4). Card et al. (2008) note that the procedure is somewhat sensitive to outliers, sometimes choosing a candidate tipping point that clearly reflects the influence of a single observation. To avoid this problem, I modify their procedure and choose the candidate tipping point via five-fold cross-validation. That is, rather than choose the candidate tipping point that minimises the in-sample R^2 of Equation (1.4), I choose the candidate tipping point that minimises the average out-of-sample R^2 of Equation (1.4).

Figure 1.3 illustrates a candidate tipping point identified in this manner for the Düsseldorf local labour market for the period 1990-1995. The vertical line marks the candidate tipping point identified by the search procedure, while the horizontal line marks the unconditional normalised native workforce growth across industries in Düsseldorf over the same period. To the left of the candidate tipping point, the normalised native workforce growth is positive and clearly larger than the unconditional average. To the right of the tipping point, it is typically negative and less than the unconditional average native workforce growth.

Figure 1.3: Growth in native workforce, Düsseldorf, 1990-1995



Notes: Growth in native workforce normalised by total workforce in base year, plotted against base year immigrant share for 3-digit industries in Düsseldorf, 1990-1995. Observations are grouped by base-year immigrant share in 2 percentage point bins, for data protection reasons. The dashed line represents fitted values from a local linear specification, estimated separately on either side of the candidate tipping point with a bandwidth of 7.5 percentage points.

In the second step, once the candidate tipping points have been identified, Card et al. propose treating the candidate tipping point as a discontinuity in the spirit of regression discontinuity (RD) designs, and estimating the jump in the conditional expectation of the outcome variable. Following this method, I will conclude that a tipping point exists if native workforce growth typically experiences a significant and negative jump at the candidate tipping point, and immigrant workforce growth experiences a positive, significant jump at the candidate tipping point. To test this hypothesis, I pool all local labour markets in a given base year and estimate the following empirical version of Equation (1.3):

$$y_{jt,t+5} = p(s_{jt} - s^*_{c(j)t}) + \delta \mathbf{1}(s_{jt} \geq s^*_{c(j)t}) + \beta X_{jt} + \alpha_{c(j)} + \varepsilon_{jt}. \quad (1.5)$$

The dependent variable, the five-year normalised change in either immigrants or natives in industry j, is assumed to evolve according to some smooth function of the distance of the base-year immigrant share from the local labour market-specific tipping point, $p(\cdot)$, with a discontinuous jump at the candidate tipping point equal to δ. I also consider as control variables X_{jt} the log of average firm size in the local industry, the log of median wages, the share of low-qualified workers and the Herfindahl-Hirschman index of concentration of the local industry, all in the base year. I follow Card et al. (2008) in modelling $p(\cdot)$ as a fourth-order polynomial in $s_t - s^*_{c(j)t}$, and I restrict my estimations to firms within 30 percentage points of their industry-year-specific tipping point.

Given that Card et al. explicitly motivate this approach by referring to the candidate tipping point as a discontinuity in the sense of RD designs, I also estimate Equation (1.3) non-parametrically via local linear regression, which is the standard in RD designs (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). As in the parametric specification, I treat the distance between an industry's base-year immigrant share and the candidate tipping point, $s_t - s^*_{c(j)t}$, as the running variable and pool all labour markets in the same base year. Again, I conclude there is a tipping point if I find a negative, significant fall in native workforce growth and a positive, significant increase in immigrant workforce growth.

It should be noted, however, that notwithstanding the language used by Card et al., the problem of testing for tipping points arguably does not correspond to an RD design. The defining feature of an RD design is that the probability of receiving some treatment jumps discontinuously when the running variable crosses a known threshold. It is not at all clear what, if anything, could correspond to the "treatment" in the present case, other than the tautologically defined treatment "having an immigrant share beyond the tipping point".

Card et al. note that using the same data to both identify the location of the candidate tipping points and test for a significant effect on workforce growth at these tipping points creates a specification search bias. This creates a risk that standard inference methods will lead us to over-reject the null hypothesis of no effect on the outcome at the tipping point. To address this bias, they randomly split their sample in two, using one subsample to search for the candidate tipping points and the other to test the effect on workforce growth. As an alternative, they also propose using the full sample to both identify the tipping point and estimate the effect on the outcome, bootstrapping their estimates to obtain standard errors, although they do not use the bootstrap approach in their main results. Subsequent work identifying tipping points in labour markets, in particular Pan (2015), has preferred bootstrap standard errors to the split-sample approach, perhaps because sample splitting requires relatively abundant data. Since my dataset is considerably smaller than that of Card et al., I follow Pan (2015) in calculating nonparametric bootstrapped standard errors for $\widehat{\delta}$ to test the null hypothesis.

1.4.2 *Threshold regression specification*

While the approach developed by Card et al. is intuitively appealing, it is ad hoc. The distribution of $\widehat{\delta}$ when following their procedure is not known, since it is a function of $s^*_{c(j)t}$, which is itself estimated. It is unclear whether either the split-sample inference conducted by Card et al. or the bootstrap inference conducted by Pan adequately deals with this problem.[5] This concern, and the observation that the tipping point is arguably not a discontinuity in the sense of an RD design, since there is no treatment variable, lead me to consider a threshold regression model as an alternative. Threshold models are conceptually appealing in this context since they are explicitly designed to deal with the situation where the conditional expectation of an outcome changes discontinuously at some unknown threshold. They do this by simultaneously estimating the location of the threshold and the effect on the outcome at the threshold as a non-linear least squares problem (Hansen, 1996, 2000). They are practically appealing since tests for the existence of a discontinuity have been developed for them and their distributional theory is well-understood (Andrews and Ploberger, 1994; Hansen, 1996, 2000).[6]

Consider the following threshold regression model, again derived from Equation (1.3):

$$y_{jt,t+5} = p(s_{jt}) + \delta \mathbf{1}(s_{jt} \geq \gamma) + \beta X_{jt} + \alpha_{c(j)} + \varepsilon_{jt}. \tag{1.6}$$

$p(s_{jt})$ is again a fourth-order polynomial, though now in the base year immigrant share. The crucial difference between equations (1.6) and (1.5) is that the tipping point is now included as γ, a parameter to be identified at the same time as δ, β, and the coefficients of $p(\cdot)$. Equation

[5] Andrews et al. (2020) have recently considered the general problem of inference conditional on an estimated breakpoint, and propose an alternative, quantile-unbiased estimator and an alternative sample-splitting approach, both of which have yet to be adopted in the literature on tipping points.

[6] More recently, related nonparametric methods have also been developed for RD designs when the location of the discontinuity in the probability of receiving treatment is unknown. Porter and Yu (2015) propose a two-step procedure where one (i) tests for the existence of a breakpoint at some unknown value of the running variable; and, should the null of no break point be rejected, (ii) estimates the location of the breakpoint as the value of the running variable for which the treatment effect is maximised. While this and related methods are interesting, they are explicitly developed within a potential outcomes framework for an RD design, which, I have argued, is not appropriate here since there is no clearly defined treatment, so I do not consider this approach further.

(1.6) is therefore nonlinear in the parameters and can be estimated by nonlinear least squares.

To test for the existence of a discontinuity, one tests the restriction that $\delta = 0$. A standard statistic for this kind of test is $T_n = \sup_{\gamma \in \Gamma} T_n(\gamma)$, where $T_n(\gamma)$ is a test (Wald, Lagrange Multiplier, F, or other kind of test) of the restriction that $\delta = 0$ when γ is treated as known. Hansen (1996) points out that the distribution of T_n is a function of γ, which is not identified under the null hypothesis that $\delta = 0$, invalidating the usual distributional theory of the test. However, he shows that a bootstrap procedure will give the correct p-values for the test. The procedure works as follows: (i) estimate (1.6) via nonlinear least squares; (ii) at each bootstrap iteration, generate a new dependent variable $y_{jt} = \widehat{\varepsilon}_{jt} z_{jt}$ where $\widehat{\varepsilon}_{jt}$ is the estimated residual and z_{jt} is a draw from a standard normal distribution; and (iii) re-estimate the model and calculate the test statistic for the generated dependent variable at each bootstrap iteration.

It is important to note the key difference between the bootstrap procedure proposed by Hansen (1996) and the one mentioned by Card et al. (2008) and used by Pan (2015). In the former case, the location of the tipping point is re-estimated in each bootstrap iteration, while in the latter case it is fixed at the value identified by the search procedure. I will show below that this has large consequences for the test that $\delta = 0$. More generally, the two-step procedure of Card et al. is related to the threshold regression approach. Equation (1.4), used to identify the candidate tipping point, is a special case of the threshold model in Equation (1.6) with the imposed assumption that the polynomial $p(\cdot)$ is of order zero, i.e. a constant function, and $\beta = 0$, i.e. the covariates do not enter the regression. Furthermore, a consistent implementation of a threshold regression would involve testing the hypothesis that $\delta = 0$ at the first step, eliminating Card et al.'s second step.

1.4.3 *Data*

The data used in this chapter come from the Institute for Employment Research of the German Federal Employment Agency. I use the Establishment History Panel (BHP), a fifty per cent sample of all establishments making social security contributions for at least one employee between 1975 and 2010. An establishment covers all production sites of a firm within the same municipality operating within the same three-digit sector. I follow standard practice when working with the BHP in indiscriminately referring to establishments as firms or establishments.

The sampling frame of the BHP includes all firms making social security contributions in West Germany since 1975, and all such firms in East Germany since 1993. I limit the sample to four five-year periods: 1990-1995, 1995-2000, 2000-2005, and 2005-2010. This allows me to investigate potential differences in tipping dynamics as immigrant flows and macroeconomic conditions change over time. I also limit myself to West Germany (excluding Berlin) since East Germany is not covered through the whole period and a large majority of Germany's immigrants live and work in the old West Germany.

I further impose the following restrictions on the observations included. I drop all local industries where fewer than 10 firms are operating, or fewer than 30 individuals are employed in the base year, to ensure that any large changes in workforce composition are unlikely to be the result of firm-specific idiosyncratic factors and to guarantee a degree of homogeneity across local industries. I also drop industries where the normalised growth in the native workforce in the period under consideration is greater than 400 per cent, since these are likely to be the result of large already-existing establishments entering the BHP after a change in establishment ID, through mergers. Finally, similar to the restriction Card et al. (2008) impose on the number of neighbourhoods per metropolitan statistical area, I restrict attention to local labour markets where at least 100 industries satisfying the above restrictions are operating, to be able to estimate local labour market-specific tipping points.[7] Focusing on large cities is less of a restriction than it may appear, since that is where most immigrants are located. After imposing these restrictions, my sample covers 60 per cent of total employment in West Germany in the BHP over the sample period, and 67 per cent of immigrant employment.

Summary statistics for the included industries are presented in Table 1.2. In the first two periods, 1990-1995 and 1995-2000, respectively 15 and 14 local labour markets covering 54 and 53 per cent of relevant BHP employment satisfy the sample definition, while in the latter two periods, 2000-2005 and 2005-2010, 24 and 25 local labour markets covering 64 and 65 per cent of relevant BHP employment satisfy the size restriction. The newly included local labour markets correspond to smaller cities and the average size of a local industry and the median real wage accordingly decline over time. One notes an increase in the average immigrant share from 1990 to 1995, consistent with the large net immigration experienced by Germany in that period, and declines thereafter.

Before evaluating dynamic patterns of workforce composition, I also report static measures of segregation in the whole of West Germany over the period under consideration, calculated from the BHP. I report two types of measures, originally proposed by Hellerstein and Neumark (2008), the index of coworker segregation and the index of effective coworker segregation. The index of coworker segregation, which measures the difference between (i) the probability that a randomly drawn coworker of an immigrant also be an immigrant; and (ii) the probability that a randomly drawn coworker of a native be an immigrant. To calculate the effective index of coworker segregation, I randomly redistribute the individuals (immigrants and natives) in my sample across firms, conditional on either firm location, industry of the firm, or both, and

[7] The local labour markets are constructed by Kropp and Schwengler (2011) from municipality commuting flows for 1993-2008 and correspond roughly to an urban core and its adjacent counties (*Kreise*).

Table 1.2: Summary statistics, local industries

	1990-1995		1995-2000		2000-2005		2005-2010	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Size	2836.54	4888.10	2897.99	4693.43	2684.20	4629.99	2649.98	4610.44
Median wage (2010 EUR)	86.24	20.35	92.10	21.86	84.88	25.90	83.00	28.48
Native growth	0.02	0.31	0.38	0.57	0.04	0.31	0.07	0.27
Immigrant growth	0.02	0.07	0.02	0.12	0.00	0.05	0.01	0.05
Workforce growth	0.04	0.35	0.41	0.65	0.04	0.35	0.09	0.30
Immigrant share	0.09	0.07	0.11	0.09	0.08	0.07	0.08	0.06
Low-education share	0.23	0.12	0.20	0.11	0.18	0.10	0.14	0.09
Share of small firms	0.69	0.21	0.70	0.19	0.70	0.19	0.71	0.18
Local industries	2001		1867		3073		3172	
Local labour markets	15		14		24		25	

Note: Reports summary statistics on either base-year characteristics or changes in workforce composition for included local industries. See main text for sample selection. The local labour markets of Bremen, Düsseldorf, Essen, Frankfurt am Main, Hamburg, Hanover, Karlsruhe, Köln, Mannheim, München, Münster, Nürnberg, Stuttgart, and Wiesbaden are included in all periods. The local labour markets of Augsburg, Freiburg im Breisgau, Heilbronn, Kassel, Kiel, Koblenz, Oldenburg, Regensburg, Saarbrücken, Siegen, and Ulm are included in at least one period.

keeping firm size equal to actual firm size, and recalculate the standard index of coworker segregation on the simulated sample. I then take the average index of coworker segregation from 30 such simulations and subtract the result from the true index of coworker segregation. The index of effective coworker segregation provides a measure of the extent to which differences between immigrants and natives in observable characteristics, in this case geographic location and industry affiliation, can explain observed patters of workplace segregation.

The obtained indexes are reported in Table 1.1. The unconditional index of coworker segregation rises somewhat over this period, from 0.15 to 0.18, as does the index of effective segregation, conditional on local labour market and three-digit industry, from 0.08 to 0.1. The figures reported here, while calculated from a different dataset, are similar to those reported in Glitz (2014), though unconditional segregation is slightly higher in my dataset, and conditional segregation is slightly lower.

1.5 RESULTS

1.5.1 *Two-step procedure*

1.5.1.1 Candidate tipping points

Table 1.3 summarises the estimated tipping points by five-year period. The location of the tipping point is assumed to be specific to each local labour market. The distribution of estimated candidate tipping points is also presented graphically in Figure 1.4. The average location of the candidate tipping point does not appear to change over different periods, with the modal tipping point around 5 per cent in each period, and the mean around 7.5 per cent. The pairwise correlation between the candidate tipping points over periods is moderately high, ranging between 0.42 and 0.72.

The distribution across local labour markets of the break in expected

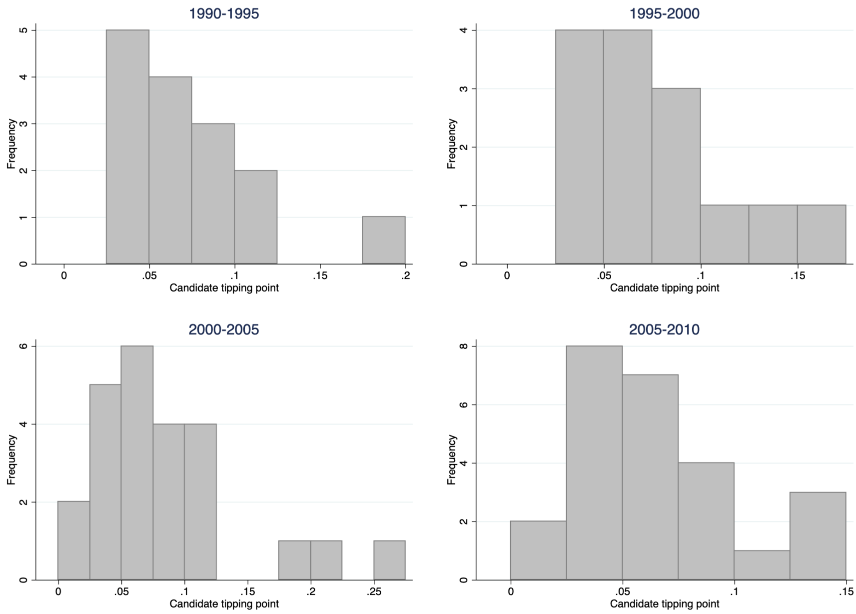
Table 1.3: Summary statistics, candidate tipping points

	1990-1995	1995-2000	2000-2005	2005-2010
Mean	0.073	0.077	0.086	0.065
St. dev.	0.043	0.040	0.062	0.038
Labour markets	15	14	24	25
Share break negative	1	0.71	0.46	0.40
Correlations				
1990-1995	1.00			
1995-2000	0.56	1.00		
2000-2005	0.76	0.50	1.00	
2005-2010	0.42	0.64	0.48	1.00

Note: Tipping points for identified candidate tipping points. # identified refers to the number of local labour markets for which there is a negative change in native workforce growth at the proposed candidate tipping point. Share significant refers to the share of identified candidate tipping points for which the drop is significant. See main text for details on the method used to calculate the tipping points. Correlations refer to pairwise correlations across base years, for local labour markets where a candidate tipping point is identified in both base years.

native growth as calculated during the search procedure, defined as $\widehat{\delta}_{c(j),t}$ in Equation (1.4), is plotted in Figure 1.5. Theoretically, $\widehat{\delta}_{c(j),t}$ should be negative, since we expect a decrease in native workforce growth as we move beyond the tipping point. However, this assumption is not imposed on the search procedure, and Figure 1.5 shows that some candidate tipping points are identified where there is a *positive* break in native workforce growth, particularly in more recent periods. In Table 1.3 I report the share of local labour markets for which the estimated break in workforce growth is negative, as expected. In 1990-1995 all breaks are negative, and the modal decline is around 15 per cent. However in 1995-2000, 2000-2005, and 2005-2010, the share of negative breaks is respectively 71 per cent, 46 per cent, and 40 per cent. This preliminary finding can already be considered *prima facie* evidence that tipping dynamics are more likely to be present in the period 1990-2000, when

Figure 1.4: Distribution of candidate tipping points



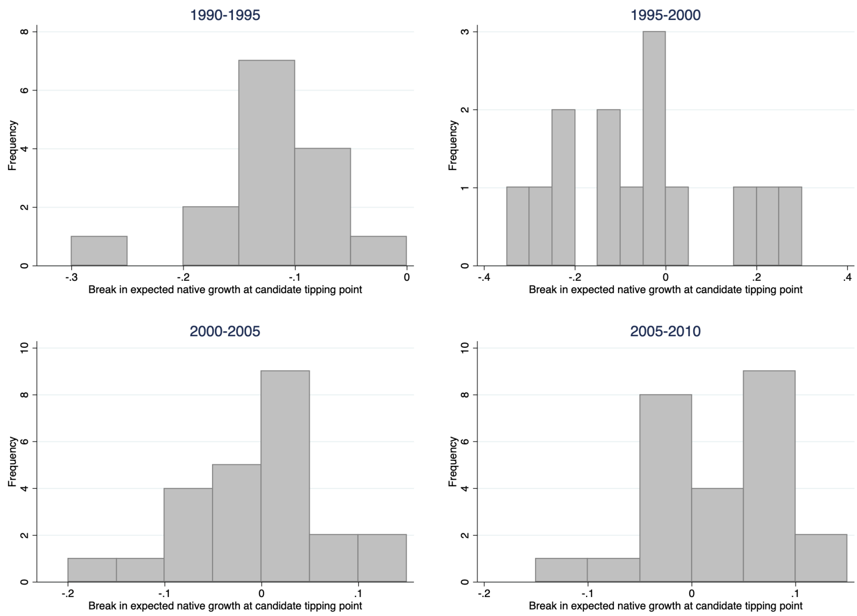
Notes: The figures plot the distribution of candidate tipping points identified by the search procedure described in Section 1.4.1. The candidate tipping points are specific to each local labour market.

immigration was higher, than in the period 2000-2010, where immigration was lower.

1.5.1.2 Discontinuities at identified candidate tipping points

I first present graphical evidence of the change in the normalised native growth rate at the tipping point in Figure 1.6. In particular, I plot fitted values calculated by estimating the pooled specification in Equation (1.5), omitting covariates and local labour market fixed effects. The graph shows clear evidence of a discontinuity in native workforce growth around the tipping point over the period 1990-1995. Normalised native

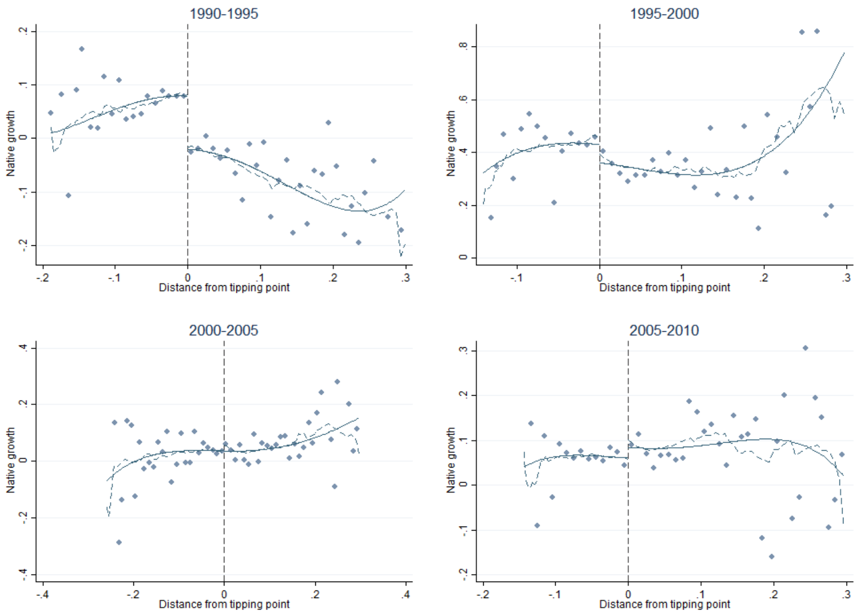
Figure 1.5: Distribution of $\widehat{\delta}$



Notes: The figures plot the distribution of $\widehat{\delta}_{c(j),t}$ the estimated change in native growth in the search procedure described in Section 1.4.1, defined in Equation (1.4). The estimated breaks are specific to each local labour market.

workforce growth drops by around 10 percentage points around the tipping point to below zero. Furthermore, a binned scatter plot of the underlying data show that there is relatively little variance from this pattern. For the period 1995-2000, the drop is similar in magnitude, around 10 percentage points, though native workforce growth remains positive to the right of the tipping point, and the binned scatter plot points to greater variance in the underlying data. For the later periods, the change is zero in 2000-2005 and even slightly positive in 2005-2010. These findings reflect the fact that for many local labour markets in these periods the search procedure identified candidate tipping points

Figure 1.6: Change in native growth at tipping point



Notes: The figures plot the change in the five-year change in the native workforce, normalised by the total base-year workforce, against the distance between the base-year immigrant share and the local labour market-specific tipping point. Local industries are binned in one percentage point bins, for data protection reasons. The solid line represents fitted values from a global parametric specification with no controls and an intercept shift; the dashed line represents fitted values from a local linear specification estimated on either side of zero, with a bandwidth of five percentage points.

where the break in native workforce growth was positive, as documented in Figure 1.5. However, the binned scatter plots show that there is some variation in the underlying data, and statistical tests will fail to reject the null of no break, positive or negative, in native workforce growth.

The observed pattern of differing effects over time is striking when considered in relation to Germany's migration history. The nineties was a period of high immigration, linked to the wars in ex-Yugoslavia and the immigration of so-called *Spätaussiedler*, typically Russian-speaking ethnic Germans from the former Soviet Union who emigrated to Germany in large numbers from the late eighties onward. In contrast, the first decade of the twentieth century was marked by several recessions and historically low rates of immigration. The fact that the relative supply of immigrants was rising strongly throughout the period 1990-2000 but much less so in 2000-2010 might explain why the observed tipping dynamics are much weaker in the latter period.

Column one of Table 1.4 presents the same information as Figure 1.6. The bootstrapped standard errors show that the positive discontinuities in the period 2000-2010 are indeed not statistically significant. However, the discontinuous drop in the period 1995-2000, while relatively large at seven percentage points, is also not significant. Column two adds covariates and local labour market fixed effects to the specifications, which do not materially alter the conclusions.

In columns three and four, I take seriously the suggestion that one think of the candidate tipping point as a regression discontinuity and estimate the change in expected native workforce growth at the discontinuity using local linear regressions, as is standard in the literature on RD designs (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The choice of bandwidth is a key parameter in these specifications, as in all local smoothing methods (Cattaneo and Vazquez-Bare, 2016). I therefore report results using both an ad hoc bandwidth equal to 0.05 in column three as a baseline, and using a mean squared error-minimising bandwidth, calculated from the data, in column four (Imbens and Kalyanaraman, 2012; Calonico et al., 2014).

The pattern of results is broadly similar to that identified using parametric methods. Using the optimal bandwidth, the effect is slightly larger in 1990-1995 at -11 percentage points but much smaller and still insignificant in 1995-2000 at -2 percentage points. The effects estimated for the period 2000-2010 are more clearly positive, and even statistically significant in 2005-2010. These positive effects in 2000-2010 are perhaps

Table 1.4: Discontinuities in workforce growth at candidate tipping points

	Global spec.		Local spec.			
	natives	natives	natives	natives	migrants	share
Discontinuity, 1990	-0.10**	-0.08**	-0.10**	-0.11**	-0.01*	0.00
	(0.03)	(0.03)	(0.03)	(0.04)	(0.01)	(0.00)
Robust p-value	–	–	0.05	0.03	0.07	0.84
Bandwidth	–	–	0.050	0.031	0.032	0.041
Local industries	1995	1995	2001	2001	2001	2001
Discontinuity, 1995	-0.07	-0.08	-0.06	-0.02	-0.00	0.00
	(0.06)	(0.06)	(0.07)	(0.09)	(0.01)	(0.00)
Robust p-value	–	–	0.65	0.98	0.90	0.83
Bandwidth	–	–	0.050	0.026	0.037	0.026
Local industries	1840	1840	1867	1867	1867	1867
Discontinuity, 2000	-0.00	0.01	0.03	0.03	0.01	0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.00)	(0.00)
Robust p-value	–	–	0.62	0.18	0.15	0.78
Bandwidth	–	–	0.050	0.049	0.041	0.039
Local industries	3059	3059	3072	3072	3072	3072
Discontinuity, 2005	0.02	0.02	0.04*	0.05*	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.03)	(0.00)	(0.00)
Robust p-value	–	–	0.01	0.04	0.61	0.93
Bandwidth	–	–	0.050	0.028	0.028	0.023
Local industries	3159	3159	3172	3172	3172	3172
Labour market FE	no	yes	no	no	no	no
Controls	no	yes	no	no	no	no

Note: Estimated discontinuity in the dependent variable at the candidate tipping point. In columns 1-4 the dependent variable is the change in the native workforce, normalised by the total workforce in the base year, in column 5 it is the normalised immigrant workforce change, in column 6 it is the change in the share of immigrants in the workforce. I report bootstrap standard errors from 1000 replications in parentheses. For the nonlinear local specifications, I also report bias-corrected robust p-values and the bandwidth used. Column 3 uses an imposed bandwidth, while columns 4-6 use data-generated MSE-optimal bandwidths. + $p<0.1$, * $p<0.05$, ** $p<0.01$.

unsurprising, given that a majority of candidate tipping points identified by the search procedure were break points where the discontinuity was positive (see Table 1.3).[8]

Previous work identifying tipping points in residential or labour markets has tended to focus on breaks in native growth. While such breaks alone are consistent with patterns of native flight or native avoidance of an industry once its immigrant share passes a certain point, theoretical work on tipping points (e.g. Becker and Murphy, 2000) has tended to focus on the composition of the group, and not only native behaviour. Column five therefore repeats the nonparametric specification, using the same candidate tipping points, only now treating immigrant workforce growth as the dependent variable. The theory outlined above predicts immigrant workforce growth should increase discontinuously at the candidate tipping point, or at least not decrease. The estimated change in immigrant growth at the candidate tipping points is typically small and insignificant, though the effect is negative and significant at the five per cent level in 1990-1995. Finally, column six combines the information on immigrant and native workforce growth by using the change in the immigrant share of the workforce over the period under study as the dependent variable in the same RD specification. In all periods, the estimated change at the candidate tipping point is less than half a percentage point, and it is never significant. This would tend to cast doubt on whether the effects on native workforce growth identified even in the period 1990-2000 can be interpreted as evidence of tipping dynamics, as opposed to a more modest form of native flight from or native avoidance of industries beyond the tipping point.

[8] In supplementary results (available on request) where I drop all labour markets where the search procedure identified candidate tipping points with a positive discontinuity, the estimated effect at the tipping point becomes negative in 2000-2010, but is small and statistically insignificant. This supplementary manual selection step is adopted by Aldén et al. (2015), however it is not justified econometrically.

1.5.1.3 Test of identifying assumption

If one does treat the test of a discontinuity in the regression function as an RD design, the key identifying assumption is that potential outcomes must change continuously around the candidate tipping point. While this assumption is not directly testable, potential outcomes being unobservable, it is typically tested either by checking that predetermined characteristics change smoothly through the threshold or by testing formally for bunching in the running variable, in this case the distance of the immigrant share from the candidate tipping point in the base year, around the threshold. Even if one is sceptical about whether the candidate tipping point corresponds to a discontinuity in the sense of an RD design, tests of the RD identifying assumption are still useful in establishing that the observed discontinuities cannot be attributed to discontinuities in other relevant local industry characteristics or to selection into having a high- or low-immigrant share by local industries.

Table 1.A.1 reports the results of local linear regressions treating different local industry characteristics in the base year as the outcome. I consider four outcomes: the share of low-skilled workers in the local industry, the log of wages, the average firm size, and the Herfindahl-Hirschman index of industry concentration. I again use MSE-optimal bandwidths derived from the data in each specification. Across the four outcomes and four time periods, I find a single significant discontinuity at the ten per cent level, in the low-skill share in 1995.

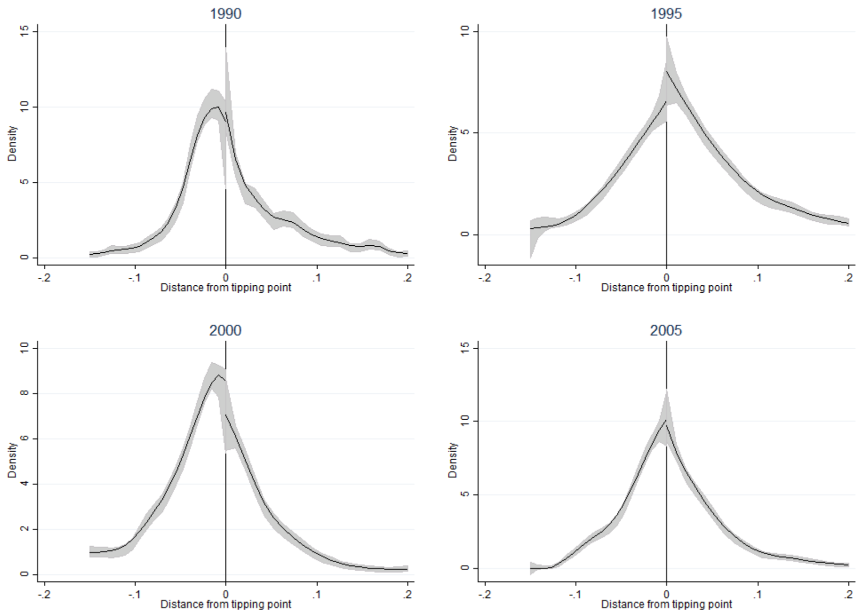
To test for bunching in the running variable, I apply the test of Cattaneo et al. (2019), which is conceptually similar to the original bunching test of McCrary (2008), but has greater power. Such tests have not been conducted in previous papers testing for the existence of tipping points using RDD-like methods, and are a rigorous complement to tests of discontinuities in other base-year characteristics. To conduct the test, I estimate the density of the running variable, the distance between the base-year immigrant share and the candidate tipping point, to the right and left of zero using local polynomials and test whether there is a significant discontinuity. Figure 1.7 plots the estimated densities and their 95 per cent confidence intervals. The confidence intervals are constructed using the bias corrected robust confidence intervals proposed by Calonico et al. (2014) and are therefore not centred on the point estimates. There does not appear to be strong visual evidence of bunching of the running variable on either side of the candidate tipping point in any period, though Table 1.A.2 shows that the formal test of discontinuities rejects the null of no discontinuity at the five per cent level in 1990.

Given that the period 1990-1995 is the one where the evidence of tipping is strongest, the existence of a discontinuity in the running variable in the base year is problematic. Given the absence of discontinuities at the five per cent level in any of the other base year outcomes considered, it is tempting to conclude that the observed bunching in the running variable is due to sampling variation. This conjecture receives some support from the fact that it is difficult to conceive of the establishments that make up a local industry, of which there are at least ten, managing to systematically collude to keep the local industry immigrant share to the right or left of an unknown cut-off point. Nevertheless, regardless of whether the bunching is due to random chance or systematic factors, its existence raises the question of whether the same factors might not also explain the observed discontinuity in native workforce growth over the period 1990-1995.

1.5.1.4 Robustness checks

One objection to the claim that there are tipping dynamics in the period 1990-1995 and not in later periods is that the set of labour markets considered is changing over time (as discussed in Section 3.3). In particular, the sample in the period 2000-2010 includes more small labour markets, where the choice of industry facing a worker may be

Figure 1.7: Distribution of base-year immigrant share



Notes: The solid line represents the estimated density of the running variable, the distance between the base-year immigrant share and the candidate tipping point in the base year, using a second-order local polynomial and MSE-optimal bandwidths, estimated separately on either side of zero. The shaded area represents a bias-corrected (asymmetric) 95 per cent confidence interval for the density, where the bias is estimated using a third-order polynomial.

more constrained, limiting the scope for tipping dynamics to arise. In Table 1.A.3, I repeat the specifications presented in this section on only those 14 local labour markets that are in my sample in all years. The total number of observations (local industries) may still change slightly from year to year, as the set of industries within each labour market is not held constant. The pattern of results and indeed the point estimates in each specification are almost identical to the full sample, although I no longer conclude that any discontinuities are significantly positive during the 2000-2010 period.

1.5.2 *Threshold regression*

In addition to the conceptual issues with the two-step approach of Card et al. (2008), noted above, the procedure does not allow us to conclusively accept or reject the existence of tipping points in the composition of local industries during the decade 1990-2000. This is an additional reason for considering threshold regressions as an alternative approach, given that such methods are well-understood and explicitly intended for the problem at hand.

1.5.2.1 Industry results

Ideally, to compare the results of threshold models with the results obtained previously, one would fit a different threshold model for each local labour market separately, thereby allowing the location of the tipping point to vary by local labour market. Unfortunately, this is not feasible, as threshold models require a relatively large number of observations to identify the threshold, since the location of the threshold is identified by observations near the threshold. Hansen (2020) recommends a sample size of $n \geq 500$, however each labour market in my sample typically does not contain more than 150 local industries. In presenting the results of a threshold model, I therefore focus on a specification where I allow for a single threshold or tipping point for all labour markets, the location of which may vary by time period. I consider specifications where I estimate a different threshold for each labour market as a robustness check.

Table 1.5 summarises my main results. Columns 1-3 use the normalised native growth as the dependent variable, successively adding the same set of base-year controls and labour market fixed effects to the specification. The location of the threshold and the effect on the outcome at the threshold are robust to these inclusions. However, native workforce growth *increases* at the identified threshold in 1995-2000 and 2000-2005, so it does not appear possible to interpret the threshold as a tipping point, at least in these years. Turning to the formal test of the discontinuity, I report bootstrapped p-value for the test of no jump in the outcome at the threshold , $\delta = 0$, following the procedure of Hansen (1996). Focusing on column 3, including controls and fixed effects, the test fails to reject the null in all specifications.

For comparison with previous work, I also report an alternative bootstrapped p-value. This p-value differs from the parametric bootstrap p-value in two respects. First it is non-parametric, constructed by repeated sampling from the empirical distribution of the observations. Second, and more importantly, it treats the location of the threshold identified by nonlinear least squares as known, and only re-estimates the remaining linear parameters of the model, including the intercept shift at the threshold $\widehat{\delta}$, at each bootstrap iteration.

Note that setting the threshold equal to the value estimated via NLS and re-estimating the linear parameters via OLS will give estimates of the linear parameters that are equal to their NLS estimates if the same data are used in both estimation procedures. To see this, observe that the NLS estimates can be obtained in two steps by (i) estimating OLS models for the range of possible threshold values in which the threshold is treated as known in each OLS estimation, and then (ii) choosing the threshold associated with the OLS model for which the sum of squared residuals is minimised. However, given that the bootstrap samples are different from the original dataset, one will obtain different estimates of the linear parameters depending on whether one estimates them via OLS with an imposed threshold, or via NLS with a simultaneously estimated threshold.

The p-values obtained from the OLS procedure are considerably smaller than those obtained by bootstrapping the full NLS procedure. In all periods the test rejects the null that $\delta = 0$ at at least the five per cent level. The reason for this large discrepancy is that the OLS bootstrap

Table 1.5: Threshold regressions with intercept shift

	natives	natives	natives	immigrants	share
1990-1995					
Tipping point	0.076	0.076	0.048	0.156	0.057
NLS p-value	0.114	0.146	0.248	0.756	0.296
OLS p-value	0.004	0.004	0.023	0.059	0.010
Discontinuous change	-0.078	-0.076	-0.076	-0.026	0.0075
Local industries	2001	2001	2001	2001	2001
1995-2000					
Tipping point	0.042	0.042	0.042	0.151	0.167
NLS p-value	0.576	0.368	0.434	0.890	0.694
OLS p-value	0.023	0.008	0.011	0.233	0.082
Discontinuous change	0.14	0.14	0.14	-0.020	-0.0077
Local industries	1867	1867	1867	1867	1867
2000-2005					
Tipping point	0.041	0.041	0.041	0.100	0.106
NLS p-value	0.376	0.526	0.680	0.068	0.022
OLS p-value	0.038	0.022	0.030	0.004	0.000
Discontinuous change	0.050	0.051	0.049	0.016	0.012
Local industries	3073	3073	3073	3073	3073
2005-2010					
Tipping point	0.059	0.062	0.062	0.101	0.101
NLS p-value	0.142	0.228	0.124	0.548	0.806
OLS p-value	0.006	0.006	0.003	0.110	0.119
Discontinuous change	-0.052	-0.048	-0.052	-0.0083	-0.0044
Local industries	3172	3172	3172	3172	3172
Controls	no	yes	yes	yes	yes
FE	no	no	yes	yes	yes

Note: In columns 1-3 the dependent variable is the normalised change in the native workforce, in column 4 it is the normalised change in the immigrant workforce, in column 5 it is the change in the immigrant share. The table reports the estimated location of the breakpoint, the tipping point, and the discontinuous change in the outcome at the tipping point. The NLS p-value is calculated from 500 parametric bootstrap iterations, allowing the breakpoint to vary, the OLS p-value is calculated from 500 nonparametric bootstrap iterations, keeping the location of the breakpoint at the value identified using the original sample. The p-values refer to the null of the test of no-discontinuities in the outcome variable.

procedure does not account for uncertainty in the location of the threshold itself. If the estimated discontinuity varies for different possible values of the threshold, then uncertainty about the location of the threshold will feed into uncertainty about the size of the discontinuity. Shutting down the first source of variability (in the location of the threshold), as the OLS bootstrap procedure does, therefore has knock-on effects on the variability of the estimate of δ.

This finding is significant, since previous studies of tipping points in the composition of neighbourhoods (Card et al., 2008; Aldén et al., 2015) or occupations (Pan, 2015) do not report measures of the precision with which the location of their candidate tipping points is estimated, nor do they account for the effect that imprecision in the estimation of the location of the tipping point will have for the estimate of the effect on the outcome at the tipping point. Indeed the nonparametric bootstrap standard errors reported by Pan (2015) are constructed in the same manner as the nonparametric bootstrap standard errors reported here. The only difference is that she fixes the threshold at the candidate tipping points estimated from the simple constant, intercept shift model of Equation (1.4), whereas the OLS p-values reported in Table 1.5 fix the threshold at the threshold obtained by estimating the full polynomial model of Equation (1.6) via NLS. The results reported here suggest it is likely that inference procedures that treat the location of the breakpoint as fixed across bootstrap samples tend to over-reject the null of no tipping points.

Turning briefly to other outcomes, I show in column four of Table 1.5 that there is no greater evidence of a tipping point when using immigrant workforce growth as the dependent variable. The discontinuity in the outcome at the estimated threshold is frequently of the wrong sign and is never significant. In column five, I treat the change in the immigrant share as the dependent variable. Here there is some evidence of a tipping point for the period 2000-2005, when immigration was low, at a base-year immigrant share of around 10 per cent. While significant, the effect is relatively small, as the expected change in the immigrant share jumps by only 1.2 percentage points.

1.5.2.2 Robustness checks

One might object that the comparison between the two-step and threshold methods does not compare apples to apples, since I allow the location of the breakpoint to vary by labour market when using the two-step procedure and keep it fixed across labour markets in the threshold regression. I therefore re-estimate the two-step procedure, searching for a single candidate tipping point by year, and re-estimate the second-step discontinuities using the same specifications as previously. The results are reported in Table 1.A.4. The results are somewhat closer to the results of the threshold specifications. The estimated discontinuities are a couple of percentage points smaller and less strongly significant in 1990-1995 than when the location of the tipping point is allowed to vary by labour market. Interestingly, the discontinuities in 2000-2010 are now negative, although still not significant. All in all, however, my conclusion remains unchanged. While point estimates in both the two-step and threshold procedures sometimes give support to the existence of tipping points, inference in the two-step procedure is over-optimistic in rejecting the null of no tipping.

As an alternative way of addressing this objection, I also estimate separate threshold models for each local labour market, using native workforce growth as the dependent variable. In only two out of the 78 labour markets and time periods considered is there a significant drop in native workforce growth at the identified threshold. However, given the sample size requirements for threshold regressions, noted above, it is difficult to conclude much from local labour market-level specifications, which use on average fewer than 150 observations per threshold regression.

Finally, I also consider the possibility that different point estimates in my threshold regressions are driven by a changing sample over time. In

Table 1.A.5 I report the results of threshold regressions using a constant set of labour markets. My conclusions are unchanged relative to the full set of local labour markets.

1.5.2.3 Firm results

Until now I have focused on tipping points in the composition of local labour markets. This was motivated by, among other reasons, an assumption that the preference spillovers that generate the tipping behaviour are primarily caused by concerns for the prestige of the industry one works in. However, one cannot rule out *a priori* that a distaste for directly interacting with immigrants in the workplace is the source of preference spillovers. In this case, the correct unit of observation would be the establishment. Furthermore, tipping in the composition of establishments does not translate mechanically to tipping in the composition of the local industry, since the composition of individual firms can tip while the immigrant share in the industry is far from the tipping point. Conversely, the immigrant share in the industry might pass the tipping point even as most firms remain far from the tipping point.

As a robustness check, to ensure that the lack of evidence of tipping points is not a result of focusing on the wrong unit of analysis, I re-estimate the threshold model treating firms as the unit of observation. This causes a practical problem, in that most firms are small (on average 70 per cent of firms in local labour markets have fewer than 10 employees), so their base-year immigrant share bunches around certain values such as 0.25, 0.33, or 0.5. Furthermore, there is selective attrition, and small firms that survive the five year period generally experience native and immigrant workforce growth that is much greater than the average. This artificially creates discontinuities in the regression of workforce growth on immigrant share around these mass points. I therefore restrict my estimates to firms employing at least 10 employees in the base year.

There is, however, a countervailing practical benefit to using firms;

since there are many more firms than industries, it is now possible to estimate a separate threshold model for each local labour market. The results are no different to the industry specifications. Of the 78 local labour markets and time periods considered, only five have a significant break in native workforce growth, and in only three is the break negative. There is, therefore, no greater evidence in favour of the existence of tipping points in the composition of firms than there is for the existence of tipping points in the composition of local industries.

1.6 CONCLUSION

Tipping-like dynamics have been identified in neighbourhood composition, school enrolments or occupational composition, and have been used to explain segregation in these different settings. This chapter considered whether such tipping points could also contribute to explaining documented patterns of segregation between immigrants and natives across workplaces.

Applying the two-stage procedure of Card et al. (2008), which first identifies tipping points and then uses regression-discontinuity methods to test for discontinuities in the evolution of the firm's workforce, this chapter has found some support for the existence of tipping points in the composition of German firms, particularly in the period 1990-2000, when immigration to Germany was high. However, the evidence is not entirely robust to the choice of estimation method and there is some evidence of bunching in the base-year immigrant share relative to the candidate tipping point.

These inconclusive results motivate me to locate candidate tipping points and test for the existence of significant discontinuities using a single unified procedure, the test for the existence of a breakpoint in a threshold regression proposed by Hansen (1996, 2000). This procedure more conclusively rejects the hypothesis of tipping points in all periods. Furthermore, comparing this method to the methods used in previous work on tipping points in labour markets (Pan, 2015) suggests that the two-step procedure where only the second step is repeated for each bootstrap sample has a tendency to under-reject the null of no tipping point. This finding is cause for some circumspection about whether there are tipping points labour markets.

Given the limited evidence of tipping points presented here, one can conclude that preference interactions are unlikely to be a leading explanation of observed workplace segregation. Future research could productively investigate what role alternative theoretical mechanisms, particularly the role of social networks in the job search process, play in explaining observed patterns of segregation.

1

APPENDIX

Appendix 1.A supplementary tables

Table 1.A.1: Discontinuities in covariates

	Low-skill share	Log wage	Average firm size	HHI
Discontinuity, 1990	0.01	-0.05	-7.39	0.00
	(0.01)	(0.03)	(7.73)	(0.01)
Robust p-value	0.43	0.06	0.33	0.74
Bandwidth	0.031	0.028	0.036	0.040
Local industries	2001	2001	2001	2001
Discontinuity, 1995	0.03^{+}	-0.04	12.06	-0.03
	(0.02)	(0.04)	(10.26)	(0.02)
Robust p-value	0.04	0.41	0.27	0.43
Bandwidth	0.023	0.026	0.028	0.027
Local industries	1867	1867	1867	1867
Discontinuity, 2000	0.02	-0.02	-0.09	-0.00
	(0.01)	(0.04)	(6.27)	(0.01)
Robust p-value	0.23	0.59	0.85	0.80
Bandwidth	0.033	0.034	0.029	0.048
Local industries	3073	3073	3073	3073
Discontinuity, 2005	0.00	0.01	-4.46	0.01
	(0.01)	(0.04)	(6.36)	(0.02)
Robust p-value	0.64	0.58	0.54	0.27
Bandwidth	0.022	0.022	0.031	0.019
Local industries	3172	3172	3172	3172

Note: Estimated discontinuity in the dependent variable at the candidate tipping point for various base year characteristics, estimated by local linear regression on either side of zero. I report bootstrap standard errors from 500 replications in parentheses. I also report bias-corrected robust p-values and the data-generated MSE-optimal bandwidth used. $^{+}$ $p<0.1$, * $p<0.05$, ** $p<0.01$.

Table 1.A.2: Test of discontinuities in the distribution of the running variable

	Test
1990-1995	
P-value	.038
Industry-cities	2001
1995-2000	
P-value	.374
Industry-cities	1867
2000-2005	
P-value	.772
Industry-cities	3073
2005-2010	
P-value	.791
Industry-cities	3172

Note: Running variable is distance between base-year immigrant share and candidate tipping point. The test is constructed by estimating the density of the running variable to the right and left of zero, using a second-order local polynomial. I report robust, bias corrected p-values for the test of no discontinuity, where the bias is estimated by fitting the density with a local third order polynomial.

Table 1.A.3: Discontinuities at candidate tipping points, constant set of labour markets

	Global spec.		Local spec.			
	natives	natives	natives	natives	migrants	share
Discontinuity, 1990	-0.11**	-0.09**	-0.11**	-0.11**	-0.01*	0.00
	(0.03)	(0.02)	(0.03)	(0.04)	(0.01)	(0.00)
Robust p-value	–	–	0.05	0.02	0.03	0.90
Bandwidth	–	–	0.050	0.034	0.034	0.041
Local industries	1895	1895	1901	1901	1901	1901
Discontinuity, 1995	-0.07	-0.08	-0.06	-0.02	-0.00	0.00
	(0.06)	(0.06)	(0.07)	(0.09)	(0.01)	(0.00)
Robust p-value	–	–	0.65	0.98	0.90	0.83
Bandwidth	–	–	0.050	0.026	0.037	0.026
Local industries	1840	1840	1867	1867	1867	1867
Discontinuity, 2000	0.01	0.02	0.04	0.05	0.01	0.00
	(0.02)	(0.02)	(0.04)	(0.04)	(0.01)	(0.00)
Robust p-value	–	–	0.58	0.14	0.17	0.38
Bandwidth	–	–	0.050	0.049	0.034	0.029
Local industries	2008	2008	2019	2019	2019	2019
Discontinuity, 2005	0.00	0.01	0.03	0.04	-0.00	-0.00
	(0.02)	(0.02)	(0.03)	(0.04)	(0.00)	(0.00)
Robust p-value	–	–	0.19	0.23	0.71	0.66
Bandwidth	–	–	0.050	0.029	0.019	0.019
Local industries	1992	1992	2002	2002	2002	2002
Labour market FE	no	yes	no	no	no	no
Controls	no	yes	no	no	no	no

Note: Estimated discontinuity in the dependent variable at the candidate tipping point. In columns 1-4 the dependent variable is the change in the native workforce, normalised by the total workforce in the base year, in column 5 it is the normalised immigrant workforce change, in column 6 it is the change in the share of immigrants in the workforce. I report bootstrap standard errors from 1000 replications in parentheses. For the nonlinear local specifications, I also report bias-corrected robust p-values and the bandwidth used. Column 3 uses an imposed bandwidth, while columns 4-6 use data-generated MSE-optimal bandwidths. $^{+}$ $p<0.1$, * $p<0.05$, ** $p<0.01$.

Table 1.A.4: Discontinuities at candidate tipping points, common tipping point by year

	Global spec.		Local spec.			
	natives	natives	natives	natives	migrants	share
Discontinuity, 1990	-0.05$^+$	-0.04	-0.03	-0.06	-0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.07)	(0.01)	(0.01)
Robust p-value	–	–	0.52	0.43	0.93	0.52
Bandwidth	–	–	0.050	0.015	0.015	0.016
Local industries	1989	1989	2001	2001	2001	2001
Discontinuity, 1995	-0.01	-0.00	-0.00	-0.02	-0.00	0.00
	(0.05)	(0.05)	(0.07)	(0.12)	(0.01)	(0.00)
Robust p-value	–	–	0.81	0.77	0.44	0.20
Bandwidth	–	–	0.050	0.021	0.019	0.022
Local industries	1842	1842	1867	1867	1867	1867
Discontinuity, 2000	-0.02	-0.01	-0.06	-0.03	-0.00	0.00
	(0.03)	(0.02)	(0.04)	(0.05)	(0.01)	(0.01)
Robust p-value	–	–	0.34	0.67	0.96	0.89
Bandwidth	–	–	0.050	0.032	0.031	0.025
Local industries	3060	3060	3073	3073	3073	3073
Discontinuity, 2005	-0.03	-0.03	-0.03	-0.01	-0.00	0.00
	(0.02)	(0.02)	(0.03)	(0.04)	(0.00)	(0.00)
Robust p-value	–	–	0.61	0.93	0.88	0.38
Bandwidth	–	–	0.050	0.011	0.013	0.013
Local industries	3147	3147	3172	3172	3172	3172
Labour market FE	no	yes	no	no	no	no
Controls	no	yes	no	no	no	no

Note: Estimated discontinuity in the dependent variable at the candidate tipping point. In columns 1-4 the dependent variable is the change in the native workforce, normalised by the total workforce in the base year, in column 5 it is the normalised immigrant workforce change, in column 6 it is the change in the share of immigrants in the workforce. I report bootstrap standard errors from 1000 replications in parentheses. For the nonlinear local specifications, I also report bias-corrected robust p-values and the bandwidth used. Column 3 uses an imposed bandwidth, while columns 4-6 use data-generated MSE-optimal bandwidths. $^+$ $p<0.1$, * $p<0.05$, ** $p<0.01$.

Table 1.A.5: Threshold regressions with intercept shift, constant set of labour markets

	natives	natives	natives	immigrants	share
1990-1995					
Tipping point	0.075	0.049	0.049	0.156	0.160
NLS p-value	0.190	0.260	0.550	0.900	0.460
OLS p-value	0.008	0.046	0.042	0.158	0.173
Discontinuous change	-0.071	-0.073	-0.073	-0.022	-0.011
Local industries	1901	1901	1901	1901	1901
1995-2000					
Tipping point	0.042	0.042	0.042	0.151	0.167
NLS p-value	0.576	0.368	0.434	0.890	0.694
OLS p-value	0.023	0.008	0.011	0.233	0.082
Discontinuous change	0.14	0.14	0.14	-0.020	-0.0077
Local industries	1867	1867	1867	1867	1867
2000-2005					
Tipping point	0.042	0.042	0.042	0.151	0.167
NLS p-value	0.560	0.330	0.420	0.890	0.710
OLS p-value	0.019	0.012	0.013	0.240	0.099
Discontinuous change	0.144	0.141	0.136	-0.020	-0.008
Local industries	2019	2019	2019	2019	2019
2005-2010					
Tipping point	0.131	0.062	0.062	0.141	0.141
NLS p-value	0.440	0.640	0.770	0.620	0.850
OLS p-value	0.024	0.044	0.047	0.153	0.140
Discontinuous change	0.067	-0.046	-0.045	0.011	0.006
Local industries	2002	2002	2002	2002	2002
Controls	no	yes	yes	yes	yes
FE	no	no	yes	yes	yes

Note: In columns 1-3 the dependent variable is the normalised change in the native workforce, in column 4 it is the normalised change in the immigrant workforce, in column 5 it is the change in the immigrant share. The table reports the estimated location of the breakpoint, the tipping point, and the discontinuous change in the outcome at the tipping point. The NLS p-value is calculated from 100 parametric bootstrap iterations, allowing the breakpoint to vary, the OLS p-value is calculated from 100 nonparametric bootstrap iterations, keeping the location of the breakpoint at the value identified using the original sample. The p-values refer to the null of the test of no-discontinuities in the outcome variable.

2

WORKPLACE SEGREGATION AND THE LABOUR MARKET OUTCOMES OF IMMIGRANTS

2.1 INTRODUCTION

A growing body of evidence has documented substantial segregation across workplaces by country of origin in developed economies (Hellerstein and Neumark, 2008; Åslund et al., 2014; Andersson et al., 2014; Glitz, 2014). Not only do immigrants tend to be segregated from natives, they also tend to be segregated from other immigrant groups. However, evidence on whether workplace segregation might contribute to persistent wage and employment gaps between immigrants and natives (Chiswick, 1978; Borjas, 1985; Lubotsky, 2007; Sarvimäki, 2011) is much scarcer. In particular, while previous work has noted a negative association between the degree of segregation of an ethnic group and average labour market outcomes of that group, there is still little evidence on the direct effect of the composition of a worker's current or past workplace on subsequent outcomes.

In this chapter, I set out to address this gap in our understanding by studying the effect of the ethnic composition of the set of coworkers in the first job held by an immigrant on the immigrant's labour market outcomes, both during and after the first job. To study this question, I use a survey of immigrants in Germany, the migrant supplement of the German Socio-Economic Panel (SOEP), that has been linked to the respondents' social security records from the Institute for Employment Research (IAB). The IAB-SOEP dataset is unique in combining systematic information on employment histories in Germany, available in the administrative data, with a wealth of survey information on immigrant's pre-migration characteristics, including pre-migration employment status, German proficiency, or the presence of social networks in Germany at migration.

Previous work has shown that more segregated groups have worse labour market outcomes on average (Åslund and Skans, 2010; Glitz, 2014) and that higher conational shares in the first job are negatively associated with individual outcomes (Ansala et al., 2021). The first contribution of this chapter is to go beyond associations and provide individual-level evidence on the effect of the conational share in the first job on subsequent outcomes. The central identification claim will be that the available set of pre-migration characteristics and information on the major decisions an immigrant makes before finding a job, such as when and where to migrate, are jointly sufficient to explain selection into first jobs with either a high or a low share of conationals.

I find that starting out in a firm with a higher conational share has a negative effect on an immigrant's probability of being employed in the longer term. A one-percentage-point increase in the initial conational share leads, on average, to a 0.16–0.18-percentage-point lower employment rate six or more years after the first job. Importantly, the employment effect is specific to the conational share, and does not exist for immigrants who do not share the immigrant's nationality, suggesting that the underlying mechanism must be specific to the conational share.

In contrast, neither the conational nor the other immigrant share appears to be significantly associated, either in the short- or long-term, with wages conditional on employment. A formal test of selection on unobservables (Oster, 2019), confirms that the estimated employment effect is unlikely to be explained by selection on unobservables. Semi-parametric estimates of the effect, using variable selection methods for treatment effects in the presence of many interactions and transformations of the control variables (Belloni et al., 2012, 2014; Chernozhukov et al., 2015), also confirm that the results are robust to alternative assumptions on the functional form.

Previous research on the earnings assimilation of migrants has shown that systematic selection into return migration with respect to realised earnings can bias estimates of the time profile of immigrant earnings in the host country (Borjas, 1985; Lubotsky, 2007). In this chapter, I am interested in the related, but distinct, question of how the time profile of employment and earnings varies with the conational share in the first job. The second contribution of this chapter is to show how systematic selection into outmigration with respect to some variable of interest, in my case the initial conational share, can bias estimates of the effect of that variable on labour market outcomes. Selection into return migration with respect to the variable of interest is a potential source of bias in all studies of how initial conditions at migration affect subsequent outcomes (e.g. Azlor et al., 2020; Battisti et al., 2018; Beaman, 2012; Damm, 2009; Edin et al., 2003; Munshi, 2003). Importantly, this form of selection is independent of the more traditional selection into the treatment on unobservables that most previous research designs were intended to minimise.

In this chapter I show how the sign of the selection bias will depend on the sign of both (i) the effect of the outcome of interest, say employment rates, on return migration; and (ii) the gross effect, i.e. without netting out any effect mediated by the outcome, of the variable of interest, the initial conational share, on return migration. Limitations in the data make it difficult to empirically assess the effect of the conational share on outmigration. However, under admittedly restrictive assumptions, I provide evidence that selection into outmigration on the initial conational share may imply that my results underestimate the full negative effect of the initial conational share on subsequent employment.

Finally, in the last part of the paper, I review the evidence for different mechanisms that might explain my finding. I find that individuals who start out in a high conational share firm are less likely to be naturalised in the long run, however differential human capital accumulation, and in particular learning German, is unlikely to explain the observed effect on employment rates. The employment effect is strongest for immigrant women, who are typically less strongly attached to the labour market than men (Sarvimäki, 2011), and highly-educated immigrants, who generally benefit more from improvements to the quality of their social networks (Edin et al., 2003). I suggest that a higher initial conational share is, therefore, likely to worsen either the rate of job offers or the distribution of wages received in the longer run, since conationals are likely to constitute a worse network than natives. However, absent more information on the quality of an individual's network or subsequent job search behaviour, it is impossible to directly test this mechanism.

There are several reasons to focus on the first job held by an immigrant upon arrival in Germany. Economists have long studied initial conditions upon an immigrant's arrival to understand how these affect an immigrant's career path. Typically, they have focused on the initial place of residence and the relationship between the size of an immigrant's ethnic group in the initial location of residence and the immigrant's subsequent labour market outcomes (Battisti et al., 2018; Beaman, 2012; Damm, 2009; Edin et al., 2003; Munshi, 2003). The switch of focus, to the initial place of work and composition of the set of coworkers, is novel. It is motivated by recent evidence that coworker networks are a more important determinant of an individual's labour market outcomes than residential networks (Eliason et al., 2019) and by evidence that an immigrant's firm identity accounts for as much as 40 per cent of the immigrant-native wage gap (Aydemir and Skuterud, 2008; Barth et al., 2012). This mirrors the more general finding that firm identity explains a substantial portion of workers' wages and wage inequality (Abowd et al., 1999; Card et al., 2013, 2018; Song et al., 2019). Here I show that a particular characteristic of the initial firm, the conational share, has persistent long-term effects on an immigrant's labour market outcomes.

Focusing on segregation in the first job an immigrant holds also has practical benefits. Job characteristics, including the conational share, are highly persistent, so characteristics of later jobs are endogenous to the characteristics of the first job. This means that characteristics of later jobs will be determined by some interaction of an (i) individual's pre-migration characteristics, (ii) their migration decisions (when and where to emigrate to), and (iii) their employment histories once in the host country (characteristics of previous jobs, duration of unemployment spells, timing of job transition etc.). The characteristics of the first job, on the other hand, will be determined by only the first two of these factors. So while the characteristics of the first job are not exogenous, they are determined by a smaller number of fixed characteristics, simplifying the identification problem somewhat.

The chapter proceeds as follows. In the following section I review predictions derived from different theories of wage determination about the effect of the composition of the set of coworkers in an immigrant's first job on an immigrant's later labour market outcomes, to structure the empirical analysis. In Section 3.3 I discuss the data used for this project. In Section 2.4 I present evidence on the association between initial workplace composition and subsequent labour market outcomes; I discuss my identifying assumption and assess to what extent the survey information I use adequately captures selection into job characteristics. In Section 2.5 I discuss different possible sources of bias: selection into the initial conational share on pre-employment characteristics, selection into return migration on the initial conational share, and model misspecification. In Section 2.6 I assess different possible mechanisms that could explain my result. Finally, section 3.6 concludes and discusses avenues for future research.

2.2 REVIEW OF RELEVANT THEORETICAL PREDICTIONS

In this section I review relevant theories that could explain any observed association between the initial conational share and labour market outcomes. I classify theories into two groups: (i) theories that predict there will be an association between the conational share when a worker starts a job and the starting wage in that job; and (ii) theories that predict there will be an association between the conational share when a worker starts a job and outcomes in later periods, such as wages and turnover—whether in the same job or in subsequent jobs—or unemployment. I call the former set of predictions contemporaneous associations and the latter long-term associations. I pay attention to whether a theory predicts that the conational share will cause an outcome, be caused by it, or simply be associated with it by sharing a common cause. I also pay particular attention to whether theories make different predictions for the conational share—the share of coworkers who are themselves immigrants from the same country of origin—and the other immigrant share—the share of coworkers who are also immigrants, but from other countries of origin.

2.2.1 *Contemporaneous associations of workplace segregation*

At a basic level, the conational share is likely to be associated with other firm characteristics. For example, it will be positively correlated with the total immigrant share and, given shares are capped at one, it is likely to be negatively correlated with the other immigrant share. It may also be negatively correlated with firm size, if immigrants tend to work in smaller, family-run firms. Inasmuch as these other firm characteristics are directly or indirectly associated with firm productivity, the conational share will also be associated with firm productivity. In models where workers' wages are an increasing function of firm productivity (e.g. Card et al., 2018; Manning, 2011), the conational and other immigrant share will then be associated with the starting wage of an immigrant.

Outside of more mechanistic relationships, the simplest form of contemporaneous association arises from a model of compensating differentials (Rosen, 1986; Sorkin, 2018). Immigrants might value the opportunity to work with conationals, and may as a consequence accept a lower wage to work in a firm where they get to work with relatively more conationals. Such compensating differentials, however, are presumably not present when working with immigrants from other countries—since immigrants may not feel much closer to immigrants from other countries than they do to natives—or at best will be strongly attenuated. In a static model of labour supply, compensating differentials will lower the wage an immigrant needs to be paid to work for a firm when the conational share is higher, although it is arguably the preference for working with conationals, not the conational share per se, that has a causal effect on wages. If a preference for working with conationals correlates negatively with unobservable individual productivity, the negative association between the conational share and the wage will likely be reinforced.

Firms might also take advantage of the networks of their employees in the hiring process to overcome information frictions. In a model where this is the case, a higher conational share has been shown to be a proxy for a newly hired immigrant having received a referral from another conational at the firm, which gives employers a more precise signal about the productivity of a match with a worker than hiring workers on the open market (Dustmann et al., 2016). Similarly, Åslund et al. (2014) argue that immigrant managers disproportionately hire other immigrants, relative to native managers, exploiting their superior information about immigrant workers.[1] In both cases, the higher conational share will be associated with more productive matches, raising the offered wage, although the association is arguably non-causal, as in the case of compensating differentials; it is the use of a referral that has a direct causal effect on wages. A higher other immigrant share is presumably not a proxy for having received a referral, or is at least a much worse proxy, so the other immigrant share will have at best a much smaller effect on wages via the use of referrals.

The conational share may also have a direct effect on the contemporaneous productivity of workers. Lazear (1999b) has noted that mixed teams likely suffer from higher communication costs, either directly, due to the absence of a common language, or indirectly, due to the absence of a shared work culture, which creates friction or misunderstandings in the workplace, lowering productivity. Empirical work has also documented specific settings where diverse workplaces are less productive, including French supermarkets (Glover et al., 2017) and Kenyan flower factories (Hjort, 2014). Lazear (1999b) argues that the existence of mixed teams implies that there must therefore be countervailing productivity gains to forming teams of workers from different cultural backgrounds. Evidence for complementarities in production between workers from different countries has been presented for specific industries, such as sport (Kahane et al., 2013), and at the aggregate level (Peri and Sparber, 2009; Ottaviano and Peri, 2012), however there is as yet no evidence that all industries enjoy productivity benefits from workforce diversity at the firm-level.[2]

The net direct effect of the conational share on worker productivity is therefore ambiguous, however in models where workers' wages are an increasing function of their marginal product, any direct effect of the

[1] Åslund et al. (2014) do not distinguish between conational and non-conational immigrants, though the point is presumably most relevant for conationals.

[2] In particular, it is not clear whether the documented aggregate complementarities derive from complementarities within production units such as firms, or through specialisation across firms.

conational share on productivity will pass through to observed wages. Other immigrants are similar to natives in this model; they impose communications costs, but may be net complements. Increasing the other immigrant share, holding the conational share constant, therefore has similarly ambiguous effects on the productivity of an immigrant. However, one might expect that the effect of the conational share will be of the opposite sign to the effect of the other immigrant share since; if communications costs dominate the gains from complementarity, the conational share will have a positive effect and the other immigrant share will have a negative effect.[3]

2.2.2 *Long-term associations of workplace segregation*

The existence of compensating differentials implies that turnover will be lower in jobs where the conational share is higher, since a higher conational share raises the reservation wage for accepting another job offer. However, compensating differentials do not imply an effect of the conational share on involuntary (from the worker's perspective) separations, and hence unemployment. Furthermore, the wage effect of compensation differentials should be constant throughout the job, the preference for working with conationals being a fixed characteristic.

If firms learn about workers' productivity on the job, then the use of referrals implies that turnover from involuntary separations (from the worker's perspective) will be lower when the conational share is higher, since employers are less likely to receive negative news about a worker's productivity (Dustmann et al., 2016; Glitz and Vejlin, 2020). By decreasing the probability of an involuntary separation, the conational share will also be negatively associated with medium-term unemployment, assuming workers spend time searching after a separation. Furthermore,

[3] One might argue that immigrants are more likely to be substitutes rather than complements, even when they don't share a country of origin; in this case there are no productivity benefits to working with more non-conational immigrants, only communications costs, so the effect of the other immigrant share will be negative.

the referral-induced wage and turnover effects of the conational share will fade with tenure, since after a time workers hired in low conational-share firms, i.e. without a referral, will only stay in the job if the firm receives relatively good news about their productivity, and adjusts the wage accordingly.

By affecting the starting wage, the initial conational share also affects the starting position of the individual on the job ladder (Burdett and Mortensen, 1998). If either the conational share or the other immigrant share are associated with a lower starting wage, for any of the reasons discussed above, they will increase turnover in the short- to medium-run, as the worker moves up the job ladder. If job offers arrive at random and are drawn from the same distribution for all workers, then the initial effect of a lower starting wage will fade out over time. However, if past wages affect subsequent wages, say because of wage bargaining where the current wage is the worker's threat point, then the effect of the starting wage may not fade out over the course of a career. A higher conational share could therefore lead to persistently lower wages if it lowers the starting wage, or higher wages if it raises the starting wage.

The use of social networks as a source of either information about job openings (Calvó-Armengol and Jackson, 2004; Boucher and Goussé, 2019) or referrals when applying for jobs (Montgomery, 1991; Galenianos, 2013; Dustmann et al., 2016) will also affect workers' wages and employment rates. Eliason et al. (2019) show that coworker networks are a particularly important determinant of labour market outcomes, more so than residential networks. Having a greater fraction of unemployed former coworkers has been shown to lower the rate of arrival of job offers for unemployed workers (Cingano and Rosolia, 2012; Glitz, 2017). It will also lower the probability of receiving a referral, since only employed workers can provide referrals, likely lowering the offered wage.

It is well-documented that immigrants have lower wages and are less likely to be employed than natives (e.g. Lubotsky, 2007; Sarvimäki, 2011). The initial conational share may, therefore, through its effects on network quality, lower the offer rate, leading to persistent differences in employment rates, and the distribution of offered wages, independently of whether the conational share affects the wage in the first job. The effect is likely to be heterogeneous by nationality; immigrants from groups with worse employment outcomes on average will be more negatively affected by starting out in a high-conational share firm. The other immigrant share will also have a negative effect on the employment and wages, although the size of the effect will depend on whether other immigrant groups have on average better employment outcomes than the worker's own group. If they have worse outcomes, other immigrants will be a worse source of information and referrals than the own group, and the negative effect of the other immigrant group will be larger in absolute value than the effect of the conational share.

Immigrants may also interact more intensively with their conational coworkers than with other types of workers, given the well-documented tendency towards homophily in the constitution of social networks (Mcpherson et al., 2001). In the terminology of Granovetter (1995), conationals might, therefore, be classified as strong ties and other workers as weak ties. Montgomery (1992) shows that if the offer rate from weak ties is higher, or the wage distribution of those offers stochastically dominates that of offers from strong ties, then increasing the share of weak ties in an individual's network will raise their reservation wage. A larger conational coworker share would therefore lower an individual's reservation wage. In particular, this effect is likely to be specific to the conational share, not the other immigrant share, since immigrants may be no more likely to interact with non-conational immigrants than with natives.[4]

The initial conational share might also affect subsequent outcomes

[4]Empirical evidence on the value of weak ties is more mixed; strong ties appear more productive in the sense that an individual is more likely to end up working with a given strong tie than a given weak tie (Gee et al., 2017b,a), however this does not imply that having more strong ties leads to higher or lower wage offers on average, as predicted by the theory. The result is also subject to selection bias, since it relies on accepted jobs, not on all job offers.

through more traditional human capital accumulation channels. Acquiring host country-specific human capital has been shown to account for a substantial portion of the convergence of immigrant wages to native wages over time (Eckstein and Weiss, 2010). Furthermore, Battisti et al. (2018) show that a higher share of conationals in the district of residence lowers the acquisition of host country-specific human capital in the longer run. They argue that this is because a larger share of conational co-residents makes job-finding easier, lowering the benefit from acquiring host country-specific human capital, though it is possible that it also raises the cost, e.g. of learnings the host country's language. A higher conational coworker share may also lower the benefit of acquiring host country-specific human capital, though the effect is likely to be attenuated relative to the co-resident conational share, since it concerns individuals who have already found a job. However, it clearly raises the cost of acquiring the host country's language, which could have a negative effect on long-term outcomes. The other immigrant share likewise probably only weakly affects the benefit of acquiring human capital, but, holding the conational share constant, it will probably raise the cost of learning the host country's language, since the worker interacts less with native speakers. This implies that the other immigrant share should also have a negative effect on long-term outcomes.

Finally, other characteristics of the initial firm may also influence the longer-term labour market outcomes of the worker. For example, starting one's career in a large firm has been argued to improve longer-term labour market outcomes (Arellano-Bover, 2020), perhaps because these firms provide more or better on-the-job training. If either the conational or other immigrant share is associated with these characteristics, they will be associated with the long-term outcomes of the worker, if these characteristics are not controlled for.

In sum, the starting conational share will have ambiguous effects on both initial wages and longer-term wages, although job-ladder models suggest both effects will be of the same sign. The other immigrant share will likely have a weaker effect, and may be of the opposite sign if there is no complementarity between immigrants of different origins. The effect of the conational share on subsequent employment, while also ambiguous, is a little easier to sign. By lowering the quality of job-finding network and the proportion of weak ties, and reducing incentives and increasing the cost of acquiring host-country specific human capital, a higher conational share is likely to increase the unemployment rate in the medium to long term. This is particularly likely to be true if one accounts for the method of finding the first job, removing any negative effect of the initial conational share, via the use of referrals, on job separations.

2.3 DATA

This project uses the IAB-SOEP Migration Sample linked to administrative data of the Institute for Employment Research (officially, the IAB-SOEP-MIG-ADIAB), which is described in detail in Brücker et al. (2013). The IAB-SOEP Migration Sample is an annual survey of individuals in Germany with a migration background (i.e. immigrants or descendants of immigrants), conducted as a supplement to the German Socio-Economic Panel (SOEP). It contains much richer information about the survey respondents than is typically available in social security data. Particularly relevant to this project, individuals who immigrated to Germany are asked about when and under what circumstances they moved to Germany, their situation before moving to Germany, their language capacity and prior knowledge of people in Germany, and how they found their first job. The survey data are then, conditional on the consent of the respondents, linked to their social security records by the Institute for Employment Research (IAB).

The construction of the dataset from the SOEP surveys and its linking to the administrative data imply an important caveat when working with the data. The only waves of the SOEP currently linked to social security data and publicly released are from 2013 and 2014. The social security data is filled in retrospectively, from 1975 to 2014. This implies that survivors, those immigrants who do not return to the home country, will be disproportionately selected into my sample. Return migrants are generally negatively selected on ability or earnings (Borjas, 1985; Lubotsky, 2007; Sarvimäki, 2011), which implies that the individuals in my sample will tend to be positively selected on unobserved labour market ability or integration potential relative to the general population of immigrants. While this type of survivor bias is common to studies of immigrants, it is nevertheless important to note that this dataset is not exempt. I will discuss the possible effects of different forms of survivor bias when interpreting my results below.

The social security data cover all periods of benefit receipt, participation in job training programmes, and employment in a job covered by the social security system. This last condition means that the self-employed and civil servants are not covered; breaks in the social security data could be indicative of unemployment or employment in one of these categories. The data are reported as notifications, which record employment or benefit receipt spells to the day. I transform the data into an annual panel, starting from the immigrant's first year of social security-covered employment. In particular, I record the fraction of days worked in the calendar year, which I refer to as an individual employment rate, the total wage earnings from social security-covered jobs in the course of the calendar year, and a dummy variable for whether the individual was employed on June 30 of the given year. Employment notifications are associated with a unique establishment identifier. Establishments correspond to all production sites of a single employer in the same municipality in the same narrowly defined industry class. I follow standard practice when working with IAB data in referring to an establishment as a firm. All establishment-level variables in the IAB data are calculated at June 30.

I restrict my attention to the subset of individuals in the linked IAB-

SOEP data who were born in a foreign country with a foreign nationality and who arrived in Germany between the ages of 15 and 64. Furthermore, individuals surveyed in the SOEP but who have never worked in a social security-covered job in Germany are by default excluded.

The final sample contains 851 individuals. I report summary statistics on the data in Table 2.1. All share variables are measured on $[0, 1]$, and wage and earnings variables are deflated to 2010 Euros. In Panel A I report time-varying information during the time following the first job in Germany. The individuals in the sample are employed a relatively high fraction of the time, particularly for immigrants, on average 74 per cent of the year. This can no doubt be attributed to positive selection into the sample, since individuals who never work a social security-covered job do not make it into the sample. In panel B I report some pre-migration time-invariant statistics. Half the sample are women and they are relatively educated on average. The average immigrant was 29 on arrival, and had a probability of 0.71 of being employed in the year before migrating; two-thirds of immigrants had support from someone in Germany at the time of migration. In panel C I report characteristics of the first job held and the firm where it was held. The average first firm is large, at 476 workers, though the distribution (not shown) is highly skewed. The social security data do not include hourly wages or hours worked, distinguishing only between full- and part-time jobs. However, notwithstanding the sample being positively selected, daily wages in the first job are on average substantially lower (43 Euros) than median daily wages in the firm (75 Euros). Just over half of my sample found their first job through contacts and they took on average 3.3 years to find that job after migrating. Finally, Table 2.2 shows the frequency of the main nationalities in my sample. The individuals in my sample are more likely to come from more recent sending countries in Eastern Europe, such as Russia, Romania, and Poland, than former guestworker-sending countries such as Turkey, Italy, and Greece.

In my results I will focus on the long-term effects of the initial

Table 2.1: Summary statistics

	Mean	St. dev.	N
Panel A			
Employment rate	0.74	0.38	9911
$P(Y > 1e4)$	0.57	0.50	9911
$P(Y > 2e4)$	0.37	0.48	9911
$P(Y > 3e4)$	0.18	0.39	9911
Annual wage earnings	21177.1	15038.7	7366
$\mathbf{1}(t \in [0,2])$	0.25	0.44	9911
$\mathbf{1}(t \in [3,5])$	0.23	0.42	9911
$\mathbf{1}(t \geq 6)$	0.52	0.50	9911
Panel B			
Woman	0.50	0.50	851
Age at migration	29.33	9.03	851
Employed before migrating	0.71	0.46	851
Low education	0.40	0.49	851
Medium education	0.32	0.47	851
High education	0.29	0.45	851
Support (family)	0.47	0.50	851
Support (friends)	0.10	0.30	851
Support (both)	0.05	0.23	851
No support	0.37	0.48	851
Panel C			
First job through contacts	0.56	0.50	851
Years until first job	3.25	3.01	851
Daily wage	42.9	34.1	851
Firm size	475.7	2236.9	851
Firm median wage	74.5	39.5	851
Firm age	13.0	10.5	851
Conat. share	0.068	0.19	851
Other mig. share	0.17	0.20	851

Note: Panel A reports time-varying summary statistics for the years since the first job, average earnings are conditional on being employed on June 30. Panel B reports summary statistics on pre-migration characteristics. Panel C reports summary statistics on the characteristics of the first job held after migration and the firm where the job was held. Wages and earnings are deflated and reported in 2010 Euros.

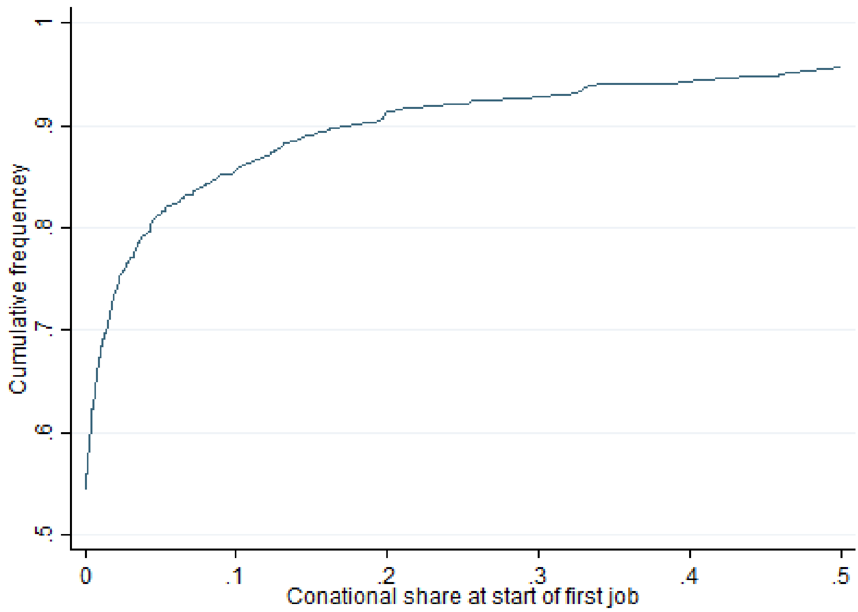
Table 2.2: Country groups

	N	Share
Russia	318	37.37
Romania	112	13.16
Poland	93	10.93
ex-Yugoslavia	71	8.34
Turkey	62	7.29
Asia	52	6.11
Italy	39	4.58
Other Europe	38	4.47
Africa	29	3.41
Greece	2*	2.**
Others	//	////
Total	851	100.00

Note: Refers to country of birth for individuals born without German nationality. The table has been censored in accordance with IAB data protection requirements.

conational share. The conational share is defined as the share of coworkers on 30 June who share the same nationality as the worker. The initial conational share is the conational share in the year of the first job subject to social security an individual holds. The average initial conational share is 0.068, while the average initial other immigrant share is 0.165. However, the distribution of the initial conational share, shown in Figure 2.1, is rather skewed. Around 55 per cent of immigrants in my sample do not have any conational coworkers at the start of their first job, while around 5 per cent of my sample start out working in a firm where more than 50 per cent of their coworkers are immigrants.

Figure 2.1: CDF of conational share in first job



Notes: Empirical CDF of the initial conational share in the first job held by an immigrant in my sample. The distribution is truncated at 50, for ease of representation.

2.4 OLS analysis

2.4.1 *Overview and identifying assumption*

In this section I present evidence on the association between the initial conational coworker share and immigrants' subsequent labour market outcomes. I will regress an outcome of interest t years after the start of i's first job, Y_{it}, on the initial conational share s_i^{own}. For now I assume the outcome follows some nonparametric time trend, $f_2(t)$, and the effect of interest, $f_1(t)$, is likewise non-constant over time, and I include a quadratic in age as relevant control variables in X_{1it}.

$$Y_{it} = f_1(t) \times s_i^{own} + f_2(t) + \Gamma_1 X_{1it} + \varepsilon_{it}. \tag{2.1}$$

The main threat to identifying the true causal effect of the initial conational share on subsequent outcomes is the possible existence of factors that (i) are pre-determined with respect to the initial conational share; and (ii) affect both the initial conational share and subsequent outcomes of interest. Obvious examples include individual preferences, such as a taste for working with conationals, and fixed characteristics, such as employability in Germany, as well as more aggregate characteristics, such as cohort effects, if the "quality" of immigrant is changing over time, nationality, or location of destination within Germany effects. There may also be individual characteristics that only indirectly affect the conational share that also directly affect subsequent outcomes. For example, the conational share is a proxy for having found a job through one's network (c.f. Dustmann et al., 2016); if less productive individuals are more likely to search for jobs through their network, this will also lead to endogeneity bias.

I address the possibility of selection on pre-employment characteristics through (i) the inclusion of fixed effects δ_j for aggregate characteristics j: nationality, year of arrival, and federal state (*Bundesland*) of first residence; and (ii) the inclusion of pre-migration characteristics available retrospectively from the SOEP, X_{2i}. Such detailed pre-migration information is not available in administrative data; its availability in the SOEP is the major advantage of using this dataset. The included characteristics are dummies for gender, being proficient in German before migration, for being employed in the year before migration, for whether the immigrant had pre-existing contacts in Germany before migrating, and for the three possible levels of education before migration, and quadratics in self-reported work experience prior to migration and age at migration.

To check how well the pre-migration characteristics and fixed effects capture selection into the first job, I regress other job and firm characteristics on the conational share, the pre-migration characteristics X_{2i}, and fixed effects δ_j, and report the coefficient on the conational share in each specification in Table 2.3. Of the job characteristics considered, only the dummy for whether the job was found through contacts is significantly associated with the conational share, conditional on included controls. The wage in the first job, in particular, is not associated with the conational share, conditional on these controls. The firm characteristics, on the other hand, are all significantly predicted by the conational share. However, this appears largely driven by the association of the conational share with establishment size. When I additionally control for log establishment size, in column two, the association between the conational share and other firm characteristics, with the exception of the other immigrant share, is substantially reduced.

I conclude from the results in Table 2.3 that the included premigration characteristics and fixed effects likely control for the determinants of the main job characteristics. Nevertheless, I will include the time taken to find a job and a dummy for whether the job was found through contacts as controls in my main specification, since these may pick up the effect of some residual confounding variable not captured by X_{2i} and the fixed effects. I also include the vector of initial firm characteristics, since these are clearly associated with the initial conational share, so that any effect of the conational share can be interpreted as holding other firm characteristics constant. I call the vector of job and firm characteristics X_{3i}.[5]

While the results presented in Table 2.3 are informative about the residual association between the conational share and job and firm characteristics, conditional on X_{2i} and δ_j, they do not allow me to conclusively rule out that there is any selection into the treatment on unobservables that also affect the outcome. I will therefore formally test my identifying assumption by testing the claim that selection on unobservables is unlikely to explain the baseline association between s_i^{own} and Y_{it}, applying the method of Oster (2019). I will present the

[5]Note that the vector X_{3i} is determined simultaneously with the initial conational share, it is not an outcome of it; it is therefore not a bad control in the sense of Angrist and Pischke (2009).

Table 2.3: Association with other firm/job characteristics

	(1)	(2)
	βs_i^{own}	βs_i^{own}
Job characteristics		
Job through contacts	0.23*	0.20*
	(0.09)	(0.10)
Years until first job	-0.00	0.00
	(0.00)	(0.00)
log(Wage)	-0.18	0.04
	(0.17)	(0.17)
Apprentice	-0.04	-0.04
	(0.02)	(0.03)
Part-time	-0.10	-0.05
	(0.08)	(0.09)
Firm characteristics		
log(Firm size)	-3.42**	
	(0.25)	
log(Median wage)	-0.66**	-0.33**
	(0.09)	(0.09)
Firm age	-11.48**	-4.35*
	(1.62)	(1.70)
Firm age^2	-310.34**	-82.24
	(49.77)	(52.05)
Other mig. share	-0.15**	-0.16**
	(0.03)	(0.04)
N	851	851

Note: The table reports the estimated coefficient on the initial conational share for a series of regressions; each row corresponds to a different dependent variable. All regressions include controls for pre-migration characteristics and fixed effects for nationality, year of migration, and location of first residence. Robust standard errors reported. + $p<0.1$, * $p<0.05$, ** $p<0.01$

structure of Oster's test in greater detail in Section 2.5.[6]

Turning from identification to estimation, to make the estimation problem more tractable, I adopt a semi-flexible approach to modelling the functions $f_1(t)$ and $f_2(t)$. Ideally, I would like to model each as a indicator variables for all values that t takes on. However, since my sample is relatively small, I group years together and instead model both functions as a set of indicator variables for being within 0-2 years of the first job, 3-5 years of the first job, or more than 6 years of the first job. The final vector of controls, X_{it}, will subsume X_{1it}, X_{2i} and X_{3i} in a single control vector, however I will introduce the three components sequentially, to assess how the estimated association changes as they are introduced. The effect of the initial share of other immigrants is allowed to vary over time, just as the effect of the conational share does. In addition to being a relevant firm characteristic that is associated with the conational share, the initial share of other immigrants will be of special interest since its effect will help to adjudicate between the different theories presented in Section 2.2. The full specification is therefore

$$\begin{aligned} Y_{it} = \sum_{g \in \{own, other\}} & s_i^g \times \mathbf{1}(t \in [0,2]) + s_i^g \times \mathbf{1}(t \in [3,5]) + s_i^g \times \mathbf{1}(t \geq 6) \\ & + \mathbf{1}(t \in [0,2]) + \mathbf{1}(t \in [3,5]) + \mathbf{1}(t \geq 6) + \Gamma X_{it} + \sum_j \delta_j + \varepsilon_{it}. \end{aligned} \tag{2.2}$$

Finally, turning from estimation to inference, in all specifications I

[6]In earlier versions of this chapter, I have also considered instrumental variables estimates of the effect of the conational share on subsequent outcomes. Asylum seekers and ethnic Germans emigrating from Eastern Europe were subject to a dispersal policy on arrival. This implies that year-on-year variations in the composition of local labour demand, and in particular the expected share of conationals for someone hired in their year of arrival, are exogenous to subsequent labour market outcomes, and can be used as an instrument for the initial conational share. However, asylum seekers and ethnic Germans are a small subset of the sample (around 200 individuals). The instrument is not strong enough to predict the conational share in such a small sample.

cluster standard errors by individual. The treatment variable, s_i^{own} is technically assigned at the level of the firm by nationality by starting year. This would be the theoretically justified level at which to cluster standard errors (Abadie et al., 2017). However, given my sample is very small, clustering observations at this level is essentially identical to clustering by individual.

2.4.2 *OLS results*

2.4.2.1 Employment rates

In Table 2.4 I report estimates of the association between the starting conational share and individual employment rates. An individual's employment rate is defined as the fraction of days they are employed in a job covered by social security in a year. Before estimating the model of dynamic effects defined in Equation (2.2), I first estimate the average effect of the initial conational share on subsequent employment, first without controls (column 1), then with controls, including the other immigrant share (column 2). I find that a one-percentage-point increase in the conational share is correlated with a 0.17-percentage-point lower employment rate, a result that is significant at the one per cent level. When including pre-migration controls, job and firm characteristics, and fixed effects, a one-percentage-point increase in the conational share is associated with a 0.11-percentage-point decrease in the employment rate, a result which is significant at the ten per cent level.

In column 3 I report estimates of the dynamic effect of initial conational share, controlling only for an individual's age and age squared. The conational share is negatively associated with subsequent employment rates at all horizons, though the effect is increasingly negative over time. In column 4 I include the pre-migration characteristics from the SOEP as controls. The effects are not statistically different from column 3 and even *increase* slightly when the pre-migration characteristics are included, suggesting individuals whose pre-migration characteristics are

Table 2.4: Relation between initial coworkers and employment

	(1)	(2)	(3)	(4)	(5)	(6)
Conat. share	-0.17**	-0.10+				
	(0.054)	(0.054)				
$\mathbf{1}(t \in [0,2]) \times$ Conat. share			-0.12**	-0.13**	-0.051	-0.00089
			(0.045)	(0.047)	(0.053)	(0.059)
$\mathbf{1}(t \in [3,5]) \times$ Conat. share			-0.19**	-0.20**	-0.14*	-0.091
			(0.067)	(0.068)	(0.065)	(0.069)
$\mathbf{1}(t \geq 6) \times$ Conat. share			-0.18*	-0.20**	-0.18**	-0.16*
			(0.079)	(0.077)	(0.069)	(0.073)
Other mig. share		-0.045				
		(0.050)				
$\mathbf{1}(t \in [0,2]) \times$ Other mig. share						-0.029
						(0.046)
$\mathbf{1}(t \in [3,5]) \times$ Other mig. share						-0.050
						(0.060)
$\mathbf{1}(t \geq 6) \times$ Other mig. share						-0.052
						(0.064)
Premigration controls	No	Yes	No	Yes	Yes	Yes
Firm controls	No	Yes	No	No	No	Yes
Job controls	No	Yes	No	No	No	Yes
Observations	9911	9911	9911	9911	9911	9911
Individuals	851	851	851	851	851	851
R^2	0.01	0.13	0.01	0.03	0.12	0.13
FE	No	Yes	No	No	Yes	Yes

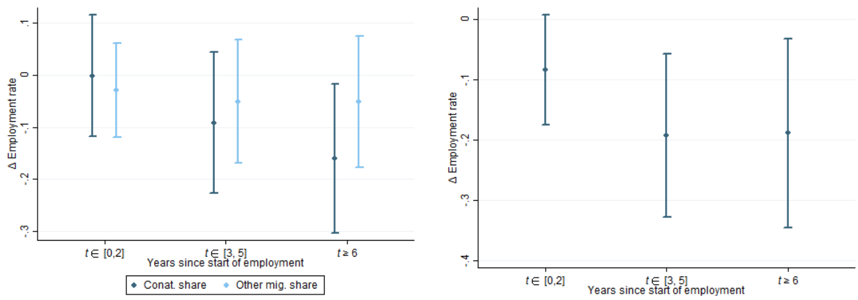
Note: OLS estimates of relationship between initial conational share and subsequent individual employment rates. The individual employment rate is the fraction of days in a year an individual is employed. The long-run coefficient is the sum of the baseline effect of the conational share (first row) and the effect at $t \geq 6$ (third row). All specifications include a quadratic in age. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01

associated with higher employment rates are more likely to work in high-conational share firms. In column 5 I add fixed effects for nationality, cohort (i.e. year of migration) and initial state of residence. The short-term association in particular decreases to -0.05 and is no longer significant; the medium- and long-term associations are less strongly reduced and remain significant.

Finally, I add other characteristics of the initial job and firm, estimating the full dynamic specification defined in Equation (2.2), and report the results in column 6. The short-term association is now indistinguishable from zero, suggesting that the short-term association between conational share and employment rates can be entirely explained by selection on observable characteristics into high-conational share firms and by the correlation of the conational share with other job and firm characteristics. The medium-term association, while still economically meaningful, is also halved by the inclusion of the full set of controls and is not significant. The longer-term association, however, is quite robust to the inclusion of all controls and fixed effects; it is reduced from -0.18 when only age is included, in column 3, to -0.16 when all controls are included, remaining significant at the five per cent level throughout. The robustness of the long-term effect as controls and fixed effects are included suggests that selection on unobservables is unlikely to account for the estimated effect; I will formally test this claim in Section 2.5.1.

In the left panel of Figure 2.2, I plot the dynamic pattern of association between the employment rate and (i) the the starting conational share; and (ii) the starting other immigrant share, estimated from the full specification including all controls and fixed effects (already presented in column 6 of Table 2.4). It is interesting to note the differing patterns between the two types of coworkers. While neither coworker share is associated with employment in the short run, given the included controls, the conational share is, as we have seen, negatively associated with long-term employment rates, while the other immigrant share is not significantly associated with employment rates. This difference is significant for at least two reasons. First, it suggests that the significant association between the conational share and subsequent employment rates cannot be explained by first jobs in firms with a higher immigrant share being of worse quality in some way that is not captured by the included controls (in particular firm size, median wage, worker starting wage and part-time status), since the association only exists with the own-group share, and not for other immigrants. Second, observing that only the conational share is associated with subsequent employment suggests that the mechanism underlying this association must be specific to the conational share.

Figure 2.2: Employment effect of composition of coworkers



Notes: Dynamic estimates of the employment effect of the initial conational share or other immigrant share. The left panel reports the coefficients from OLS estimates, the right panel reports semi-parametric estimates using the post-regularisation method of Chernozhukov et al. (2015). The post-regularisation estimates do not necessarily retain the other immigrant share as regressors, so these are not reported. 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

To put the magnitude of the long-term association into context, Glitz (2014) finds that the average employed immigrant in Germany in 2008 had 18 percentage points more conational coworkers than would be expected under a random allocation of workers, or 13 percentage points after partialling out the effects of region of residence, gender, education, and industry. The unemployment rate of the foreign-born in Germany at the time was 12.3 per cent, 5.8 percentage points higher than the unemployment rate of the native-born (OECD, 2020). Scaling the long-term effect of the conational share by average segregation translates to a $0.16 \times 18 = 2.9$ percentage point lower employment rate, or 2.1 percentage points if observable characteristics are partialled out of the measure of segregation. The magnitude of the long-term association between the initial conational share and unemployment is therefore large relative to the difference in employment rates between immigrants and natives in Germany.

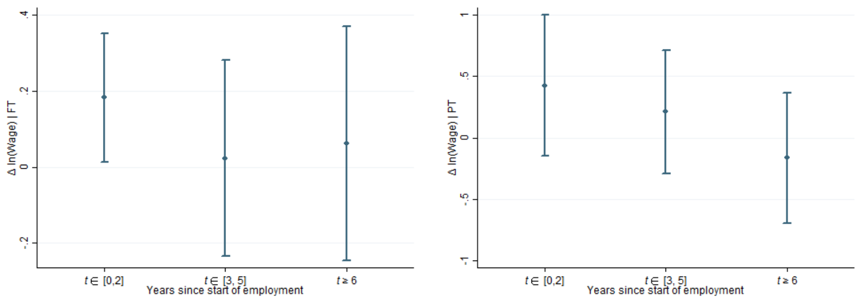
While the results in Table 2.4 show that immigrants are less likely to be in a job subject to social security in subsequent years if their first job is in a high conational share firm, it is not possible to assert based on this result alone that the individual is more likely to be unemployed. In Table 2..9 I explore other measures of an individual's labour force status as outcomes, including my full set of controls X_{it} and fixed effects in all specifications. In columns 1-4 I consider measures drawn from the administrative data: share of days in a year of benefit receipt, share of days as a registered job seeker, share of days in a job training program, and a dummy for being out of the social security system altogether. Only the last of these variables is (positively) associated with the conational share. Individuals out of the social security system might be genuinely unemployed, or they might be in self-employment or civil servants. In columns 5-6 I draw on the SOEP survey, which for the years 2013 and 2014 asks if individuals are employed and, if so, in what activity. In particular, I define dummy variables equal to one for individuals who report either self-employment or working as a civil servant. While the sample is much smaller, there is no economically or statistically significant long-run association between these variables and the initial conational share. I therefore conclude that a higher initial conational share is associated with an increased probability of an individual dropping out of the labour force in the longer term.

2.4.2.2 Wage earnings

In Table 2.5 I repeat the full specification, including fixed effects and controls for pre-migration, initial job and initial firm characteristics, for different measures of wages and earnings. The social security data only include daily wages and an indicator for part-time status. In column 1 I therefore estimate the association between the initial conational share and average daily earnings, defined, for individuals who work at least one day during the year, as total earnings subject to social security in a year divided by total number of days worked, deflated to 2010 values. There does not appear to be a significant relationship between the initial conational share and average daily wages, conditional on employment. The estimated magnitude is also small; a one-percentage-point increase in the initial conational share increases earnings 0.2 log points. To account for any possible effect of the initial conational share on average daily hours worked, I repeat the estimation for respectively full- and part-time workers. Part-time status and the daily wage are here measured on June 30 of a given year, the results are reported in Figure 2.3 and in columns 2 and 3 of Table 2.5. While the initial conational share is positively associated with daily wages of full-time workers in the short-term, there is no longer-term association. For part-time workers there is no association at any horizon.

While the evidence reported in columns 1-3 of Table 2.5 suggests there is little significant association between the initial conational share and earnings, these estimates will suffer from selection bias. Individuals who are employed, whether full-time or part-time, in spite of having a high conational share in their first job are potentially positively selected on unobserved employability relative to other immigrants, introducing a conditional-on-positive selection bias (Angrist and Pischke, 2009). This kind of selection would likely bias the estimated association between the initial conational share and potential subsequent earnings upward relative to the true association in the full, unobservable, population. As

Figure 2.3: OLS estimates of earnings effect



Notes: OLS estimates of the dynamic effect of the initial conational share on daily wages for full-time workers (left panel) and part-time workers (right panel). 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

Table 2.5: Relation between initial coworkers and earnings

	(1) ln(Avg. wage)	(2) ln(Wage) $\mid$ FT	(3) ln(Wage) $\mid$ PT	(4) $P(Y > 0)$	(5) $P(Y > 1e4)$	(6) $P(Y > 2e4)$	(7) $> P(Y > 3e4)$
$\mathbf{1}(t \in [0, 2]) \times$ Conat. share	0.20	0.18*	0.43	0.084+	0.036	0.091	0.14**
	(0.13)	(0.086)	(0.29)	(0.048)	(0.062)	(0.061)	(0.049)
$\mathbf{1}(t \in [3, 5]) \times$ Conat. share	-0.014	0.024	0.21	-0.026	-0.073	-0.091	0.088
	(0.14)	(0.13)	(0.26)	(0.070)	(0.080)	(0.075)	(0.062)
$\mathbf{1}(t \geq 6) \times$ Conat. share	-0.22	0.062	-0.16	-0.12	-0.17+	-0.17+	-0.025
	(0.17)	(0.16)	(0.27)	(0.073)	(0.086)	(0.090)	(0.078)
$\mathbf{1}(t \in [0, 2]) \times$ Other mig. share	0.11	0.20*	0.25	0.0023	0.059	-0.018	-0.0076
	(0.12)	(0.080)	(0.22)	(0.035)	(0.067)	(0.054)	(0.042)
$\mathbf{1}(t \in [3, 5]) \times$ Other mig. share	0.10	0.070	-0.43	-0.056	0.039	-0.032	0.016
	(0.13)	(0.090)	(0.31)	(0.058)	(0.075)	(0.066)	(0.050)
$\mathbf{1}(t \geq 6) \times$ Other mig. share	0.046	0.039	-0.46+	-0.042	-0.032	0.033	0.032
	(0.15)	(0.11)	(0.27)	(0.059)	(0.078)	(0.077)	(0.069)
Observations	8422	4923	2390	9911	9911	9911	9911
Individuals	851	693	536	851	851	851	851
R^2	0.34	0.43	0.26	0.12	0.21	0.30	0.32

Note: OLS estimates of relationship between initial conational share and subsequent earnings. Y refers to annual labour earnings covered by social security. The regression for average earnings in column 1 is estimated conditional on an individual being employed in a job covered by social security at least one day during the year, daily wages in columns 2 and 3 are measured on June 30 of the relevant year. All coefficients are estimated using the specification defined in Equation (2.2), standard errors are clustered by individual. + p<0.1, * p<0.05, ** p<0.01

such, it is not possible to conclude whether the true effect of the initial conational share on wages is zero, or negative but biased toward zero when conditioning on individuals being employed.

To avoid conditional-on-positive selection bias, in columns 4-7 of Table 2.5 I use the full sample and regress a dummy for annual earnings being above a series of cutoffs on the initial conational share and the full set of controls and fixed effects. This approach is conceptually similar to a quantile regression, however the interpretation of regression coefficients is more straightforward. The cutoffs I consider are 0, 10,000, 20,000, and 30,000 Euros. An increased conational share does not appear to uniformly shift the distribution of earnings. There is some evidence of a positive short-run association between the initial conational share and earnings; in particular a one-percentage-point increase in the initial conational share increases the probability of earning more than 30,000 Euros by a statistically significant 0.14 percentage points. Given the absence of employment effects at this horizon, and given the firm and pre-migration characteristics controlled for, including whether the job was found through a contact, this positive association suggests that immigrants do earning higher wages when working with more conationals, perhaps because they are more productive. Note that there is again no effect for immigrants from other countries. The initial conational share is negatively associated with long-term earnings, though given the magnitude of this effect is broadly in line with the negative long-term employment effects documented above, it is not possible to conclude from this that there is any strong evidence of a long-term wage effect of the initial conational share.

The finding of a clear negative effect of the starting conational share on long-term employment and at best only a transient and, if anything, positive wage effect is consistent with the finding that the total earnings gap between immigrants and natives is mostly due to differences in employment, not wages conditional on employment (Sarvimäki, 2011). It is also broadly in line with the theoretical mechanisms reviewed in

Section 2.2, where I argued that different theories made conflicting predictions about the short- and long-term wage effects, but a more clear prediction of a negative longer-term employment effect of the initial conational share. Having established that there is a significant negative association between the initial conational share and employment rates, I now turn to assessing possible sources of bias that could explain this finding.

2.5 POTENTIAL SOURCES OF BIAS

2.5.1 *Selection on unobservables into the treatment*

The central identification claim of this chapter is that the extensive set of controls before and at migration included in my main specification, made possible by the information gathered in the SOEP, allow me to plausibly control for unobserved pre-employment characteristics that might lead to selection into a first job with a higher or lower conational share. While the robustness of the long-term effect of the conational share on employment to the inclusion of controls and fixed effects provides some support for this claim, it does not formally rule out the possibility that the effect could be explained by selection on unobservables. Here I formally test whether selection on unobservables is likely to explain the observed effect of the initial conational share on employment rates and wages.

2.5.1.1 Overview

Intuitively, the test that I will apply involves comparing two sets of estimates: (i) a non-causal association between a variable of interest and an outcome that might be at least partially explained by selection on some variable; and (ii) an association between the same variable of interest and outcome, this time controlling for variables that are thought to measure the characteristics on which selection takes place. Because selection is thought to explain a part of the uncontrolled association, one expects the coefficient of interest to change between (i) and (ii). However, if there truly is an underlying causal effect, this change should not be "too big". Just how big is too big is determined by the change in R^2 between the two regressions. Altonji et al. (2005) use this insight to develop an estimator of the ratio between (a) the (unobserved) covariance between the variable of interest and the unobserved confounders and (b) the covariance between the variable of interest and the observed confounders that would make the true causal effect of the variable of interest zero.

To construct their test, Altonji et al. (2005) assume that the R^2 of the regression would be one if all confounders were included. Oster (2019) observes that this is unduly restrictive if there is an idiosyncratic component to the outcome of interest or if variables are measured with error. She therefore develops a generalised version of the test that allows the maximum R^2 to be less than one. Again, as in the test of Altonji et al. (2005), the central observation is that movements in the estimated treatment effect alone as covariates are included in the model are not informative about the possible extent of remaining selection on unobservables unless they are scaled by movements in the R^2. Intuitively, only treatment effects that are robust to the inclusion of covariates that actually explain the outcome should be labelled robust. The output of Oster's test is again an estimate of the ratio between (a) the covariance between the unobserved confounders and the treatment variable; and (b) the covariance between the treatment variable and the included confounders that would be consistent with the true treatment effect being zero. I refer to this estimated ratio as Oster's δ.

The maximum possible value of the regression R^2, i.e. when all observed and unobserved confounders are included in the regression, R_{max}, is a key ingredient in estimating Oster's δ. Oster (2019) suggests that $R_{max} = \min\{1.3 \times \widetilde{R}, 1\}$, where $\widetilde{R}$ is the R^2 from the long regression including all controls, is a reliable benchmark. Reviewing evidence from randomised experiments, where selection on unobservables can be ruled out *a priori* if randomisation succeeded, she finds that using this value of R_{max} would lead the researcher to conclude that 10 per cent of experimental results were due to selection on unobservables. In reviewing a selection of articles from well-known journals, she finds that around 50 per cent of published effects would be explained by selection on unobservables using this standard.

The calculation of Oster's δ is only defined for a scalar treatment variable. I therefore report the estimated δ both for selection on unobservables in the time-invariant specification, and separately for the each time horizon in the dynamic specification. To estimate the δ, the researcher must also specify the set of controls that are intended to capture selection into the treatment. I am principally concerned about individuals selecting into high- or low-conational share first jobs based on unobservable characteristics that are predetermined relative to their taking up those jobs. I argue that my included pre-migration characteristics, drawn from the SOEP, and characteristics at migration, captured by my fixed effects for year of migration, location of arrival, and nationality, are good controls for unobservable predetermined individual characteristics. However, I argued previously that initial firm and job characteristics may also capture some residual selection. I therefore only include age and age squared in the short regression (and, in the dynamic specification, the interactions of the initial conational share and years since migration that are not being tested for selection on unobservables).[7]

2.5.1.2 Results

Table 2.6 reports the estimated values of Oster's δ for my employment specification in column 1. In the static specifications, in the first row, the

[7] My conclusions about the likelihood of selection on unobservables do not change if I focus only on my pre-migration characteristics and fixed effects and either include firm and job controls in both regressions, or exclude them from the calculation of δ entirely.

δ for the employment regression is 1.72. Following Altonji et al. (2005), Oster (2019) suggests that 1 is a reasonable cutoff for declaring results robust to selection on unobservables, since $\delta < 1$, implies that the true treatment effect could be zero even if there is less selection into treatment based on the unobservables than on the observables. The value in the static specification is above the cutoff; for the observed association to be explained by selection on unobservables, these unobservables would have to be almost two times as strongly correlated with the initial conational share than the observables are. The pattern of estimates of δ for the dynamic effects clearly mirrors the pattern of point estimates. The short-term δ is close to zero, the effect is not at all robust to selection on unobservables, while the medium- and long-term effects are increasingly robust, the value of δ in these two cases is 2.16 and 4.42. I can therefore conclude with a high degree of confidence that the long-term effect in particular is robust to selection on unobservables.

Table 2.6: Estimates of Oster's δ

	$P(E=1)$	ln(Wage) | FT	ln(Wage) | PT
Average effect	1.72	-1.18	-0.23
$t \in [0,2]$	0.019	-1.48	-1.72
$t \in [3,5]$	2.16	-0.21	-1.05
$t \geq 6$	4.42	-5.51	0.72
Π	1.3	1.3	1.3
R^2_{max}	0.18	0.56	0.34
N	9911	4923	2391

Note: Estimates of Oster's δ, the ratio of the selection on the observable to the selection on the unobservables implied by model estimates and an assumed value of R^2_{max}. I assume $R^2_{max} = \min\{\Pi\widetilde{R}, 1\}$, where $\widetilde{R}$ is the R^2 of the long regression, including all controls. For details, see main text.

In columns 2 and 3 I report estimates of δ for the effect of the conawtional share on wages conditional on either full- or part-time employment. In this case, the estimated δ is typically negative. This occurs because the included covariates cause the estimated effect to increase in magnitude; the unobservables would therefore have to push the estimated effect in the other direction for the true effect to be zero. Unlike the test of Altonji et al. (2005), Oster's δ is well-defined in the case where the included covariates increase the estimated effect; the effect is now declared to be robust to selection on unobservables if $\delta < -1$. The wage effects are not particularly robust to selection on unobservables; the static δ for full-time workers is -1.18, which is marginally robust, while $\delta = -0.23$ for the static specification for part-time workers. When looking at the values of δ in the dynamic specifications, the short-term effect appears most robust to selection on unobservables, as $\delta < -1$ in both cases, although $\delta = -5.51$ in the long-term for full-time workers. However, given the long-term wage effect is zero, it is not clear that such a large negative value of δ is meaningful.[8]

The results of these tests for selection on unobservables strengthen the claim that the associational effect of the initial coworker share on subsequent employment, estimated in Section 2.4.2.1, likely captures the true causal effect. In particular, they provide formal support for the claim that the rich set of pre-migration characteristics, including pre-migration employment, work experience, proficiency in German, having contacts in Germany before migrating, and fixed effects capturing differences across cohorts, nationalities, or location of arrival in Germany, adequately capture selection into high- or low-conational share firms.

2.5.2 *Selection on the treatment into return migration*

Having formally established that selection on unobservables into high-conational share firms is unlikely to explain the effects estimated in

[8] Bevis et al. (2020) claim that the Stata command `psacalc` which estimates δ can sometimes be unreliable when $\delta < 0$. It is possible that this is what occurs in this case, given that the long-term wage effect is zero.

Section 2.4, I now formally address the possible effects of sample selection bias. I have already noted that my sample is made up of survivors, immigrants who were still in Germany in 2013 and 2014 in order to be interviewed. It is generally accepted that return migrants had worse labour market outcomes, summarised by earnings, before returning than immigrants who stay (Borjas, 1985; Lubotsky, 2007; Sarvimäki, 2011). This tells us that earnings have a negative effect on return migration, or that return migration and earnings share some common unobservable cause—return migrants might be intrinsically less productive individuals—either of which can bias estimates of the rate of earnings convergence of immigrants to natives over time (Abramitzky et al., 2014). However, when studying the effect of some initial condition, whether the ethnic network at migration or, as in my case, the conational share in the first job, on subsequent labour market outcomes, the sign of the selection bias will depend not only on the effect of earnings on return migration, but also on the effect of the initial conational share on return migration.

2.5.2.1 The sign of the bias under no confounding

To focus on intuition and to emphasise the fact that the bias induced by selective return migration is independent of the bias induced by selection into treatment on unobservables, I derive the sign of the selection bias under the simplifying assumption that (i) the initial conational share, S is randomly assigned; and (ii) there are no systematic determinants of subsequent employment rates Y besides S. Furthermore, assume that the conational share is either low or high, i.e. $S \in \{0,1\}$. Assuming the effect of S on Y is linear, the structural equation for Y is simply:

$$Y = a + \beta S + \varepsilon_Y. \tag{2.3}$$

The structural error term ε_Y is mean-zero[9] and independent of S, since there is no confounding. To model selection, I assume that latent utility

[9] Furthermore, we must have $\varepsilon_Y \in [-a, 1-(a+\beta)]$, since $Y \in [0,1]$

C^* is a linear function of S, Y, and a mean-zero structural error term:

$$C^* = \alpha_S S + \alpha_Y Y + \varepsilon_{C^*}, \tag{2.4}$$

where $\alpha_i \in \mathbb{R}$, $i \in \{Y, S\}$. An individual is assumed to return migrate, $C = 1$, if latent utility is below some fixed threshold:

$$C(S,Y) = \begin{cases} 1 & \text{if } C^* < K, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Equation (2.5) captures the fact that C is endogenously determined by both S and Y. The sign of α_i, $i \in \{Y, S\}$, encodes hypothetically testable assumptions about the effect of the observable variables Y and S on C. I now show how the selection bias from conditioning the analysis on $C = 0$ depends on the signs of α_S, α_Y, and β. Since the structural equation is linear and S is assumed to be randomly assigned, the true parameter of interest, β, can be defined as

$$\beta = \frac{\mathrm{Cov}(Y,S)}{\mathrm{Var}(S)} \tag{2.6}$$

Since we only observe individuals with $C = 0$, however, the OLS estimand on this restricted sample is

$$\begin{aligned} \widehat{\beta} &= \frac{\mathrm{Cov}(S,Y|C=0)}{\mathrm{Var}(S|C=0)} \\ &= \beta + \frac{\mathrm{Cov}(S,\varepsilon_Y|C=0)}{\mathrm{Var}(S|C=0)} \\ &= \beta + \frac{\mathrm{Cov}(S,\varepsilon_Y|C^* \geq K)}{\mathrm{Var}(S|C^* \geq K)} \end{aligned} \tag{2.7}$$

The sign of the bias induced by conditioning on the endogenous variable C will therefore depend on the sign of the conditional covariance of S and ε_Y, since the conditional variance of S is positive. Note that $\mathrm{Cov}(S, \varepsilon_Y) = 0$ in the full sample by assumption, but not in the restricted sample of non-return migrants. The sign of the conditional covariance can be calculated as

$$\begin{aligned}
\mathrm{Cov}(S,\varepsilon_Y|C^* \geq K) \\
&= \mathrm{E}[S\varepsilon_Y|C^* \geq K] - \mathrm{E}[S|C^* \geq K]\mathrm{E}[\varepsilon_Y|C^* \geq K] \\
&= \mathrm{E}[\varepsilon_Y|C^* \geq K, S=1]\Pr(S=1|C^* \geq K) \qquad (2.8) \\
&\quad - \mathrm{E}[S|C^* \geq K]\mathrm{E}[\varepsilon_Y|C^* \geq K] \\
&= \{\mathrm{E}[\varepsilon_Y|C^* \geq K, S=1] - \mathrm{E}[\varepsilon_Y|C^* \geq K]\}\Pr(S=1|C^* \geq K), \qquad (2.9)
\end{aligned}$$

where the second equality follows from the law of iterated expectations and the third from the fact that S is a Bernoulli random variable, so its expectation is the probability that $S = 1$. The sign of the conditional covariance will depend on the sign of the difference of the two conditional expectations in parentheses in Equation (2.9), $\mathrm{E}[\varepsilon_Y|\cdot]$. Note, however, that ε_Y is a mean-zero random variable and that its distribution is truncated when calculating the expectations $\mathrm{E}[\varepsilon_Y|\cdot]$. The sign of the conditional expectations will therefore depend on whether the right or the left tail of the distribution is truncated. Furthermore, the difference between the expectations will depend on which distribution is more severely truncated. The truncation condition $C^* \geq K$ can be re-written

$$\alpha_Y \varepsilon_Y \geq K - (\alpha_S + \alpha_Y \beta)S - \alpha_Y a - \varepsilon_{C^*}, \qquad (2.10)$$

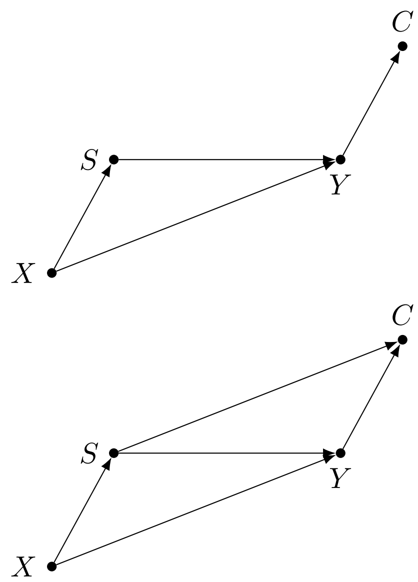
This inequality makes clear how the sign of the bias of $\widehat{\beta}$ with respect to β will depend on (i) the total effect of employment on return migration, captured by α_Y; and (ii) the total effect of the conational share on return migration, that is without netting out the part of the effect that is mediated by employment, i.e. $\alpha_S + \alpha_Y \beta$. Intuitively, the sign of α_Y determines whether the distribution of ε_Y is left- or right-truncated, and the sign of $\alpha_S + \alpha_Y \beta$ determines whether the distribution is more or less severely truncated when $S = 1$. If both α_Y and $\alpha_S + \alpha_Y \beta$ are of the same sign, the bias will be negative, while if α_Y and $\alpha_S + \alpha_Y \beta$ are of opposite signs, the bias will be positive. A formal proof of this claim is in Appendix 2.A.1.

An interesting special case arises when the true effect of interest $\beta = 0$. Now the gross effect of the conational share on return migration is simply the direct effect, α_S. In this case, if α_Y and α_S are of the same sign, then $\widehat{\beta} < 0$, while if they are of opposite signs, then $\widehat{\beta} > 0$. Therefore, if the estimated $\widehat{\beta} < 0$ and one has reason to believe that α_Y and α_S are of opposite signs, then the observed association cannot be entirely explained by selection into return migration; it must be that $\beta < 0$.

2.5.2.2 The sign of the bias in the presence of confounding

In Appendix 2.A.2 I consider a more general model of selection where S and Y may share common causes X and $S \in [0, 1]$. The sign of the bias in this case now depends non-linearly on more parameters, obscuring the nature of the selection problem created by conditioning the analysis on the endogenous variable C, which can be more clearly shown using a causal graph. Figure 2.4 depicts the relationship between the observable variables in two possible causal graphs. Time flows from left to right in these graphs, and the presence of a directed edge between two variables indicates the existence of a causal effect. Longer paths connecting two variables will create supplementary associations between them, unless either (i) a variable on the path is conditioned on, e.g. included as a control in a regression; or (ii) the path includes a so-called collider variable, a variable that is caused by both a variable that precedes it and a variable that succeeds it along the path of interest, and that collider variable is *not* conditioned on. For example, in the top panel of Figure 2.4, the causal effect $S \to Y$ is the object of interest, however there is a supplementary non-causal association between S and Y via their common causes, the confounders X, i.e. along the path $S \leftarrow X \to Y$. We therefore include X as a vector of control variables in the regression, to remove this non-causal association from the total association between S and Y, leaving only the true causal effect $S \to Y$.[10].

Figure 2.4: Possible structural relationships as graphs



Notes: Possible causal structures relating initial coworker share S to subsequent employment Y, their common (observed) causes X, and return migration C. In the bottom panel, C is a collider along the path $S \to C \leftarrow Y$.

[10]Causal graphs were originally developed in computer science and epidemiology and are complementary to approaches using potential outcomes. The conditions on a graph for identifying the causal effect of one variable on another are, under mild assumptions, equivalent to the (conditional) independence assumption required to identify a causal effect defined as a difference in potential outcomes. The general advantage of the graphical approach to causal relations is that it is possible to make statements about, and think through possible sources of bias only in terms of (potentially) observable variables, rather than in terms of unobservable counterfactual variables. See Pearl (2009) for a canonical presentation of causal graphs, Hernán and Robins (2020) or Morgan and Winship (2014) for discussions of the relationship between potential outcomes and causal graphs, and Imbens (2020) for a discussion of their applicability in economics. I consider the potential outcomes formulation of the same selection problem in Appendix 2.A.3

Focussing on selection into return migration, in the top graph of Figure 2.4, the initial conational share S has no direct effect on return migration C, only an indirect effect via subsequent earnings, Y. Conditioning the analysis on $C = 0$ therefore does not create any new associations between S and Y, which are only connected via the paths $S \rightarrow Y$ and $S \leftarrow X \rightarrow Y$; controlling for X allows us to estimate the causal effect of S on Y for the subpopulation with $C = 0$. In the second graph, however, S has a direct effect on C. C is now a collider variable along the path $S \rightarrow C \leftarrow Y$; conditioning the analysis on $C = 0$ creates a supplementary, non-causal association between S and Y along this path, even when the vector of controls X is included in the regression. This graphical presentation makes clear that bias induced by selection into return migration is independent of whether all common causes of S and Y have been conditioned on and depends on the existence of an effect both of Y on C and of S on C.

2.5.2.3 Evidence of selection on the treatment into return migration

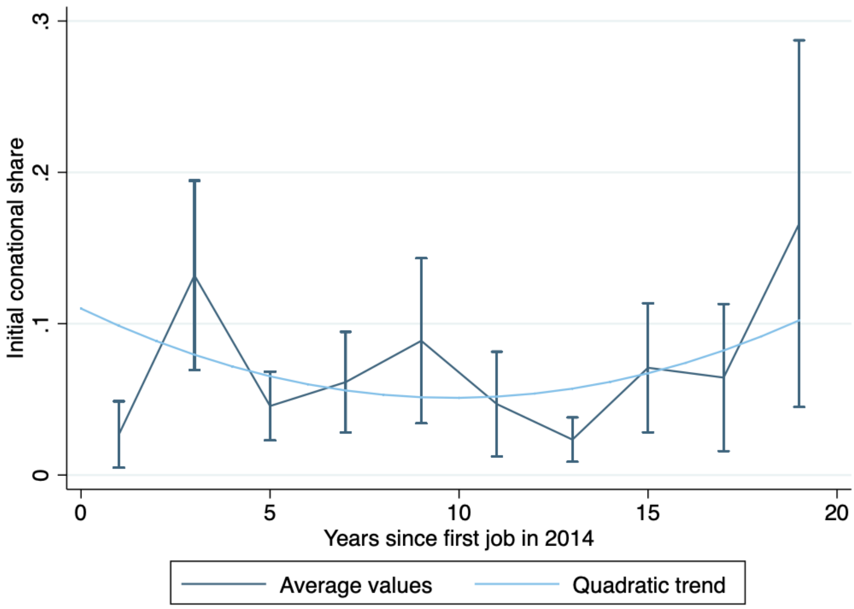
In Section 2.4.2 I estimated that $\widehat{\beta} < 0$. Assuming the selection bias is not so strong as to change the sign of the effect, one could conclude that $\beta < 0$. There is good evidence that lower earnings and employment make an individual more likely to re-emigrate (Lubotsky, 2007; Sarvimäki, 2011; Abramitzky et al., 2014), implying that $\alpha_Y > 0$, i.e. the opportunity cost effect dominates the effect of any target savings behaviour. All that remains to be determined is the sign of α_S, the association between S and C^* after partialling out Y. In a dataset that does not contain any return migrants, at least at the time of observation, it is not possible to show direct evidence of the sign of α_S. Nevertheless, it is possible to provide indirect empirical evidence on the relationship between the initial conational share and selection into outmigration by comparing cohorts that were first employed in Germany more or less recently in the year the individuals were sampled, in my case, 2014. If there are no year-of-first-employment effects, i.e. the starting conational share is the same for all newly employed cohorts, and a higher initial conational share induces greater rates of return migration, then the initial conational share will be higher in more recently employed cohorts, as fewer of the individuals who started out in a high-conational share firm have yet re-emigrated.[11]

In Figure 2.5 I show the unconditional relationship between time since first employment in Germany and the initial conational share in 2014, binning observations by year of first employment and plotting a quadratic trend in time since first employment in 2014. There is some evidence of the initial conational share decreasing and then plateauing with time since first employment, suggesting that α_S might be negative. While there is some evidence of the initial conational share increasing again for individuals who have been in Germany more than 15 years, this may be simply the result of observing fewer individuals who have been in Germany that long.

The coefficients of the quadratic trend, reported in column 1 of Table 2..10 are significant at the ten per cent level. However, the causal graph presented in Figure 2.4 makes clear that the unconditional association between S and C plotted in Figure 2.5 is the combination of the direct causal effect of interest, $S \to C$, an indirect causal effect $S \to Y \to C$, and the non-causal association $S \leftarrow X \to Y \to C$. To be able to infer the sign of α_S, one needs to control for the effect of Y on return migration, blocking both the indirect causal path and the non-causal association. OLS estimates of the time trend controlling for the average individual employment rate between first employment year and 2014 are reported in column 2 of Table 2..10. The time trend is now statistically insignificant although the magnitudes are still economically relevant: an individual in her first year of employment in 2014 had a coworker share

[11] If, on the other hand, the initial conational share has a trend over time, it will not be possible to identify both the cohort effects and the effect of the conational share on return migration using data only on stayers in a given year.

Figure 2.5: Starting conational share over time



Notes: Differences in initial conational share for immigrants who first worked in Germany t years ago in 2014. Average values are grouped in two-year bins for data protection reasons.

on average 3 percentage points higher than an individual in her fifth year of employment who had not yet return migrated.

While the evidence presented here relies on the strong assumption that there are no cohort effects in the initial conational share and is estimated on a small sample, it nevertheless suggests that $\alpha_S < 0$ is not an unreasonable assumption. In this case, if it is true that $\alpha_Y > 0$ and $\beta < 0$, then the selection bias will be positive and $\beta < \widehat{\beta}$.

2.5.3 *Model misspecification bias*

2.5.3.1 Overview

While I have shown that selection on the conational share into return migration and selection on unobservables into high-conational share firms are unlikely to explain the employment results, it is possible that my estimates nevertheless suffer from some model misspecification bias. In particular, I have imposed restrictive assumptions on the form of the regression function to make it tractable, such as the assumption that continuous variables affect the outcome linearly (or sometimes quadratically).

To check that my estimates are robust to more flexible functional forms without overfitting my relatively small sample, I would like to allow for a wide set of interactions between my control variables and only retain ones that are truly relevant. Traditional dimension-reduction methods of penalised estimation, such as the Least Absolute Shrinkage and Selection Operator (LASSO) treat all regressors as equal, and may not retain my regressor of interest, S in the set of included predictors of the outcome Y. Furthermore, the LASSO and related methods are not intended to estimate the marginal effect of any one variable on the outcome Y, so even if S is retained as a regressor by the LASSO, it is incorrect to interpret the estimated coefficient on S as an estimate of the true marginal effect of S on Y.

For this reason, methods for applying the LASSO to causal and structural models and conducting inference on a set of linear parameters of interest break the set of predictors of Y into two groups: one low-dimensional group of regressors of interest (in this case S and its interactions with years since first employment, though here I focus on S for expositional ease) and one high-dimensional set of nuisance regressors, whose inclusion is necessary to guarantee that the structural model is correctly specified, X. Elements of X are then chosen by regressing Y and S one-by-one on the set X using the LASSO. The marginal effect of

S on Y can then be estimated by calculating the residual of the LASSO regression of Y on X and regressing this on the residual of the LASSO regression of S on X, an approach known as post-regularisation (Belloni et al., 2013; Chernozhukov et al., 2015).[12]

2.5.3.2 Results

I consider the following set of control variables X: (i) orthogonalised fifth-degree polynomials in age, pre-migration experience, age at migration, log wages in first job, and log firm size, firm log median wages, and firm age, all in the first job; (ii) dummy variables for each nationality group, year of migration, federal state in which first located, and education group, as well as dummy variables for being employed and for being proficient in German pre-migration, for having a first job that was part-time or an apprenticeship, for gender, for having support from contacts in Germany when moving, and for finding the first job through contacts; (iii) all one-way interactions for the complete set of dummy variables; (iv) all one-way interactions between the dummy variables and the terms of the fifth-degree polynomials; and (v) dummy variables for years since migration and their interactions with the initial other immigrant share.[13] In total, this makes for 1220 control variables in my high-dimensional nuisance regressor set.

By design, if two regressors are highly correlated, the LASSO will usually only retain one of them, which cannot be interpreted to mean that only the retained variable matters for the outcome. Nevertheless, it can be instructive to consider the set of retained variables as a check

[12] Note the conceptual similarity of this approach to the Frisch-Waugh-Lovell theorem, where one regresses the residual of a regression of Y on a low-dimensional X on the residual from regressing T on X.

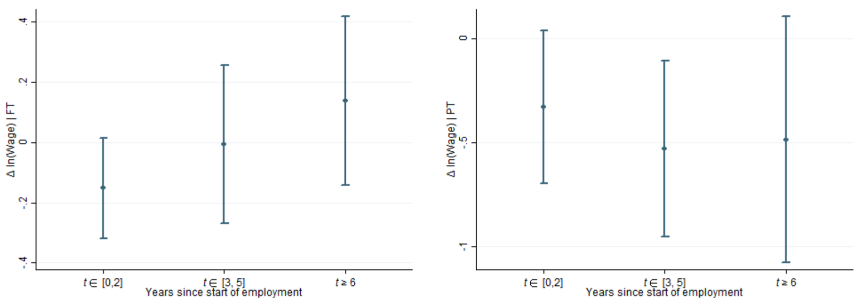
[13] This implies that I retain the assumption that the effect of the included covariates is constant over time, with the potential exception of the other immigrant share. In results available on request, I check that my results are robust to including interactions of all dummies and polynomial terms with the years since migration dummies. The estimated effect of the initial conational share remains negative and significant, however interpreting the larger set of retained covariates is less straightforward.

of the researcher's priors. For both the employment rate and earnings conditional on being employed, the LASSO retains the set of time since first job dummies, the interaction of the share of other immigrants with $\mathbf{1}(t \in [3,5])$, a dummy for Romanian nationality, the interaction of the Romanian dummy with dummies for part-time, first job through contacts, gender, and the full set of education dummies, and linear terms in age of first establishment and log first establishment size. The employment specification also includes in particular a dummy for having contacts in Germany at migration and pre-migration German proficiency, while the earnings specifications include, in particular, linear terms for the log starting wage and log median wage in the first firm and the interaction of quadratic terms for the same variables with a dummy for being high-educated.

It is instructive and perhaps reassuring to consider that measures of employability, such as pre-migration German proficiency and having a pre-existing network of contacts in Germany matter (positively) for subsequent employment, but not wages conditional on employment, while measures of the quality of the first job, in particular starting wage and median firm wage, are important predictors of subsequent wages, but not of subsequent employment. Conversely, the differential effect of several factors for Romanians, the second-largest group in my sample, was not necessarily expected *a priori*.

In the right panel of Figure 2.2 I plot my semi-parametric estimate of the dynamic employment effect over time. The effect is if anything stronger than the parametric estimate, presented in the left panel. Already in the medium term a one-percentage-point increase in the conational share is associated with a highly statistically significant 0.19-percentage-point decline in the employment rate, a decline that is also present in the longer term. There is even modest evidence of a decrease in employment rates in the short-term. I compare semi-parametric estimates of the effect on daily earnings, conditional on being employed either part-time or full-time, in Figure 2.6. These show that the modest positive earnings effect estimated by OLS for full-time workers is not robust to a more flexible functional form. Indeed for both full- and part-time workers there is evidence of a small negative short-term effect, which appears to persist into the medium-term for part-time workers. It bears emphasising, however, that the semi-parametric earnings estimates, which condition on being employed, still suffer from selection bias; it is not possible to conclude from these estimates that there is a negative causal effect. The semi-parametric estimates are repeated for convenience and standard errors are reported in Table 2..11.

Figure 2.6: Post-double selection estimates of earnings effect



Notes: Post-regularisation estimates of the dynamic earnings effect of the initial conational share on daily wages for full-time workers (left panel) and part-time workers (right panel). 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

To summarise, in this section I have shown that the employment effect estimated in Section 2.4.2.1 is robust to selection on unobservables, selective return migration, and more flexible regression specifications. There does not, however, appear to be strong evidence that the earnings effect is different from zero.

2.6 DISCUSSION OF POSSIBLE MECHANISMS

Having established that the effect of the conational share on subsequent employment is robust to different possible sources of bias, in this section I explore what evidence there is for the different theories outlined in Section 2.2. In particular, I suggested there that the main mechanisms through which the initial conational share might worsen the employment rate of immigrants were by reducing their incentives to acquire host country-relevant human capital, particularly language skills, or by worsening the quality of their social network, thereby reducing the arrival rate of job offers.

I review evidence for different possible explanations for the estimated effect in Table 2.7 where I regress alternative outcomes on the conational share, conditional on the full set of controls and fixed effects defined in Equation (2.2). In column 1 I test the persistence of the conational share, conditional on the full set of controls. These estimates necessarily condition on individuals being employed. While there is some long-term persistence, this is not very high; a one-percentage-point increase in the conational share is associated with a 0.14-percentage-point increase in the conational share for employed workers six or more years later, and is not associated with the other immigrant share at any horizon. This suggests that the estimated employment effect is unlikely to be explained by certain types of individuals always working in high-conational share firms, where it might be harder for them to find jobs. In column 2 I test whether the conational share is associated with turnover, i.e. leaving a job, conditional on controls and fixed effects. There is a small short-term association, but no longer-term association, suggesting the employment effect is not explained by individuals in high-conational share firms finding it harder to hold onto a job in the long run.

The SOEP survey asks respondents about their current knowledge of German. I can therefore directly test whether the conational share has an effect on individuals' learning German. In column 3 I regress a

Table 2.7: Relation between initial coworkers and measures of social integration

	(1) Conat. share	(2) Job separation	(3) Proficiency	(4) Naturalised	(5) Visited home	(6) Feel German	(7) Foreign friends
$\mathbf{1}(t \in [0,2]) \times$ Conat. share	0.71**	0.11*	-0.49**	-0.12	-0.37*	0.040	0.32+
	(0.046)	(0.050)	(0.16)	(0.094)	(0.16)	(0.20)	(0.18)
$\mathbf{1}(t \in [3,5]) \times$ Conat. share	0.39**	0.021	-0.30+	-0.15	-0.18	-0.15	0.042
	(0.082)	(0.055)	(0.17)	(0.11)	(0.15)	(0.14)	(0.32)
$\mathbf{1}(t \geq 6) \times$ Conat. share	0.14*	0.014	-0.087	-0.35**	0.026	-0.17*	0.023
	(0.057)	(0.040)	(0.088)	(0.073)	(0.085)	(0.079)	(0.11)
$\mathbf{1}(t \in [0,2]) \times$ Other mig. share	0.0038	0.093*	-0.45*	-0.33*	-0.44**	-0.20	-0.21
	(0.025)	(0.043)	(0.18)	(0.13)	(0.17)	(0.16)	(0.27)
$\mathbf{1}(t \in [3,5]) \times$ Other mig. share	-0.033	0.045	-0.22+	-0.19	-0.22	-0.11	0.24
	(0.025)	(0.051)	(0.13)	(0.15)	(0.14)	(0.15)	(0.19)
$\mathbf{1}(t \geq 6) \times$ Other mig. share	0.0076	0.066+	-0.085	-0.055	-0.0075	0.013	0.18
	(0.035)	(0.038)	(0.081)	(0.082)	(0.083)	(0.079)	(0.11)
Observations	7560	9911	1663	1652	1629	1646	820
Individuals	851	851	838	836	838	836	820
R^2	0.29	0.04	0.28	0.32	0.25	0.13	0.15

Note: OLS estimates of the relationship between the initial coworker share and measures of social integration, drawn from the 2013 and 2014 SOEP survey. The long-run coefficient is the sum of the baseline effect of the conational share (first row) and the effect at $t \geq 6$ (third row). All specifications include a quadratic in age, controls for pre-migration characteristics, and first job and firm characteristics as well as the full set of fixed effects defined in the text. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01

indicator for being proficient in German at the time of the survey on the conational share. The sample is restricted to the years 2013 and 2014, when the SOEP was conducted, meaning my estimates are likely to be less precise. A-one-percentage point increase in the initial conational share is associated with a 0.3–0.5-percentage-point lower probability of being proficient in German in the first five years of the immigrant's time in Germany, however not in the longer term.[14] Working with more conationals does, therefore, appear to slow down an individual's learning German. However, the effect is not persistent, suggesting that it is perhaps unlikely to explain the longer term reduction in employment caused by the initial conational share. Corroborating this claim, the association of the initial conational share and of the other immigrant share with subsequent German proficiency are almost identical. This suggests that even if lowered German proficiency in the medium term were a mechanism by which the initial conational share lowered employment in the longer term, it could not be the only, or even the primary mechanism by which this happens. Otherwise, one would observe, contrary to the fact, that the other immigrant share has a similar negative association with subsequent employment.

Turning to alternative measures of cultural assimilation, also recorded in the SOEP, which might be proxies for having acquired more "soft" Germany-specific skills or cultural knowledge, in columns 4-7 of Table 2.7 I evaluate the effect of the conational share on the probability of being naturalised, on having visited the home country in the past two years, on reporting feeling "completely or mostly" German, and on reporting that the majority of one's friends are foreign (only available in 2013). A one-percentage-point higher conational share is associated with a 0.35-percentage-point lower probability of being naturalised in the longer run, and a 0.17-percentage-point lower probability of reporting feeling

[14] Note that the set of controls includes a dummy for having been proficient in German *before* migrating, derived from a separate question in the SOEP survey questionnaire.

German. These effects are not present in the shorter run, nor are they present for the other conational share. However, it is entirely possible that they are a consequence, not a cause of a reduced attachment to the labour market. Finally, a lower conational share is associated with a higher probability of having a majority of foreign friends in the short run, but not in the longer run.

In the absence of strong evidence that differential human capital accumulation mediates the effect of the conational share on subsequent employment, there is relatively greater support for network-based theories that suggest that a higher conational share will slow immigrants' progress up the job ladder. Without observing the use of networks or job search methods to find later jobs, I cannot directly show that these explain my findings, however I do show indirect supporting evidence.

Assessing patterns of heterogeneity in the effects of the conational share provides a measure of support for the claim that worse social networks explain my findings. In columns 1 and 2 of Table 2.8 I re-estimate my main specification separately for men and women. The effect is clearly strongest for women, for whom a one-percentage-point increase in the conational share lowers the long-term employment probability by 0.33 percentage points. Immigrant women are typically less attached to the labour force than immigrant men, and have lower employment rates (see e.g. Sarvimäki, 2011). It seems reasonable that they would therefore be more likely to drop out of the labour force entirely if their job offer rate declines, or the distribution of offered wages deteriorates. This, and the fact that the conational share should affect the incentives to learn German equally for men and women, provides some support for network-based explanations of the negative employment effect.

There is also an interesting pattern of heterogeneity by pre-migration education level, reported in columns 3-5. In particular, medium- and highly-educated immigrants, those with at least an apprenticeship qualification, are more susceptible to the negative effects of starting out with a low conational share. However, for highly educated individuals,

Table 2.8: Heterogeneity of employment effect

	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}(t \in [0,2]) \times$ Conat. share	-0.053	0.071	-0.039	0.0046	0.077
	(0.098)	(0.074)	(0.091)	(0.12)	(0.13)
$\mathbf{1}(t \in [3,5]) \times$ Conat. share	-0.16	-0.048	-0.050	-0.14	-0.035
	(0.11)	(0.085)	(0.090)	(0.13)	(0.17)
$\mathbf{1}(t \geq 6) \times$ Conat. share	-0.33**	-0.082	-0.021	-0.23+	-0.30+
	(0.12)	(0.080)	(0.082)	(0.13)	(0.17)
$\mathbf{1}(t \in [0,2]) \times$ Other mig. share	-0.12+	0.066	-0.073	-0.029	-0.072
	(0.071)	(0.065)	(0.082)	(0.084)	(0.083)
$\mathbf{1}(t \in [3,5]) \times$ Other mig. share	-0.13	0.026	-0.036	0.041	-0.25*
	(0.091)	(0.080)	(0.097)	(0.100)	(0.11)
$\mathbf{1}(t \geq 6) \times$ Other mig. share	-0.15	0.025	-0.036	0.029	-0.31*
	(0.11)	(0.079)	(0.10)	(0.093)	(0.13)
Observations	4613	5298	4311	3160	2440
Individuals	428	423	338	270	243
R^2	0.16	0.20	0.18	0.22	0.26
Sample	Women	Men	Low	Med	High

Note: OLS estimates of the relationship between initial conational share and subsequent individual employment rates. The individual employment rate is the fraction of days in a year an individual is employed. Columns 1 and 2 report results conditional on gender, columns 3-5 conditional on the pre-migration educational attainment being either lower than apprenticeship, an apprenticeship, or higher than apprenticeship. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01

the effect is also present for the other immigrant share. These patterns are consistent with previous evidence showing that highly educated immigrants benefit more from improvements in the quality of their ethnic network (Edin et al., 2003). These results suggest that the negative effect of having many conational coworkers and, for the highly educated, many immigrant coworkers of other nationalities, may be more likely to stem from the reduced average quality of the total set of coworkers, rather than from changing the proportion of strong (conational) versus weak (native or other immigrant) ties.

2.7 Conclusion

In this chapter I have shown that starting one's career in an establishment with a high share of conationals has negative long-term effects for an immigrant's labour market outcomes. This is in contrast to the literature on initial residential conditions for newly arrived immigrants, where a high share of conationals in an immigrant's location of residence, by expanding the size of an individual's network, is generally thought to have positive effects on an immigrant's labour market outcomes.

One common feature of the existing results on the effects of initial residential conditions is that they potentially suffer from selection bias due to differential selection into return migration based on the treatment of interest. This chapter provides the first formal treatment, to my knowledge, of the sign of the bias this is likely to create for estimates of the effect of initial conditions. The results contained in this chapter could be productively used in future research to empirically assess the sign of the different components of the bias in these settings. Such an exercise would require a dataset that can identify future return migrants and non-return migrants, something that is not possible with the present dataset.

I suggest that starting in a high conational share firm may worsen the job offer arrival rate, since conationals are a worse source of information about the labour market. However, without observing subsequent characteristics of one's coworkers, such as their employment rate, or the job offer rate, it is impossible to test this hypothesis directly. Future work would ideally test this mechanism directly, by looking, for example, at how the effect of the conational share varies with the average quality of conationals in the location of residence.

2

APPENDIX

Table 2..9: Relation between initial coworker share and other labour market outcomes

	(1) Benefit receipt	(2) Jobseeker	(3) In training	(4) Not in IEB	(5) Self-employed	(6) Civil servant
$\mathbf{1}(t \in [0,2]) \times$ Conat. share	-0.015	-0.024	-0.043**	-0.050	0.15	-0.0028
	(0.056)	(0.060)	(0.013)	(0.035)	(0.16)	(0.013)
$\mathbf{1}(t \in [3,5]) \times$ Conat. share	-0.016	-0.016	-0.043**	0.033	0.38*	-0.031
	(0.061)	(0.062)	(0.016)	(0.048)	(0.18)	(0.020)
$\mathbf{1}(t \geq 6) \times$ Conat. share	0.059	0.070	-0.024+	0.12+	0.079	0.0075
	(0.071)	(0.075)	(0.014)	(0.069)	(0.062)	(0.0079)
$\mathbf{1}(t \in [0,2]) \times$ Other mig. share	-0.036	-0.10*	-0.016	0.0037	0.023	0.014
	(0.049)	(0.050)	(0.021)	(0.024)	(0.035)	(0.014)
$\mathbf{1}(t \in [3,5]) \times$ Other mig. share	-0.038	-0.026	-0.017	0.013	-0.054	-0.012
	(0.055)	(0.056)	(0.021)	(0.032)	(0.041)	(0.013)
$\mathbf{1}(t \geq 6) \times$ Other mig. share	0.11+	0.12+	0.0021	-0.047	-0.0056	-0.0075
	(0.063)	(0.065)	(0.016)	(0.030)	(0.033)	(0.0060)
Observations	9911	9911	9911	9911	1494	1494
Individuals	851	851	851	851	837	837
R^2	0.15	0.16	0.06	0.08	0.15	0.08

Note: Benefit receipt is a dummy for receiving earnings replacement benefits or unemployment benefits, defined respectively under Social Code Book (SGB) III and SGB II, in the course of the year. Jobseeker is a dummy for being registered as a job seeker with an employment agency. In training is a dummy for participating in a federal or state active labour market policy measure. Not in IEB is a dummy for no social security data being available in a given year. Self-employed and Civil servant are self-reported dummy variables from the SOEP survey, available in 2013-14. All specifications follow Equation (2.2), standard errors clustered by individual. + $p<0.1$, * $p<0.05$, ** $p<0.01$.

Table 2..10: Relationship between initial conational share and return migration

	(1)	(2)
	Conat. share	Conat. share
t	-0.012^{+}	-0.0092
	(0.0064)	(0.0068)
$t \times t$	0.00061^{+}	0.00052
	(0.00035)	(0.00036)
Average Employment rate		-0.057^{+}
		(0.033)
Constant	0.11**	0.13**
	(0.026)	(0.028)
Observations	782	782
R^2	0.005	0.010

Note: Evidence of a relationship between the conational share and time since first job, via the relationship between time since first job, t, and the conational share. Robust standard errors reported. + p<0.1, * p<0.05, ** p<0.01

Table 2..11: Semi-parametric estimates

	(1) P(Employed)	(2) ln(Wage) | FT	(3) ln(Wage) | PT
$\mathbf{1}(t \in [0,2]) \times$ Conat. share	-0.084^{+}	-0.15^{+}	-0.33^{+}
	(0.046)	(0.085)	(0.19)
$\mathbf{1}(t \in [3,5]) \times$ Conat. share	-0.19**	-0.0068	-0.53*
	(0.069)	(0.13)	(0.22)
$\mathbf{1}(t \geq 6) \times$ Conat. share	-0.19*	0.14	-0.49
	(0.080)	(0.14)	(0.30)
Observations	9911	4923	2391
Individuals	851	693	537

Note: Semi-parametric estimates of the effect of initial conational share on subsequent outcomes; control variables and interactions chosen via post-regularisation (Chernozhukov et al., 2015), see main text for details. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01

Appendix 2.A The bias induced by selective return migration

2.A.1 Proof of the sign of the bias under no confounding

I claim that if $\alpha_Y > 0$, the bias will be of the opposite sign to $\alpha_S + \alpha_Y \beta$, while is $\alpha_Y < 0$, the bias will be of the same sign as $\alpha_S = \alpha_Y \beta$. To see this, note that if $\alpha_Y > 0$, the condition $C^* \geq K$ truncates the left tail of the distribution of ε_Y; the expectations in Equation (2.9) will be positive. Furthermore, if $\alpha_S + \alpha_Y \beta > 0$, then the supplementary condition $S = 1$ truncates the distribution less severely than when the condition is not imposed, since $S \in \{0, 1\}$. As a result, we will have

$$\mathrm{E}[\varepsilon_Y | C^* \geq K, S = 1] < \mathrm{E}[\varepsilon_Y | C^* \geq K] \tag{2.11}$$

and the bias will be negative. If, on the other hand, $\alpha_Y < 0$, the right tail of the distribution is truncated and the expectations in Equation (2.9) are negative. If $\alpha_S + \alpha_Y \beta > 0$, the supplementary condition $S = 1$ again means the distribution is less severely truncated, implying now that

$$\mathrm{E}[\varepsilon_Y | C^* \geq K, S = 1] > \mathrm{E}[\varepsilon_Y | C^* \geq K] \tag{2.12}$$

and the bias will be positive.

2.A.2 The sign of the bias with confounding

I derive an expression for the bias of OLS estimates in the presence of selection on the treatment variable S into return migration $C = 0$ in the presence of covariates X. I continue to assume that the variables are linear functions of each other, however I drop the assumption that S is Bernoulli, and allow $S \in [0, 1]$. The structural representation of the graph in the bottom panel of Figure 2.4 is therefore

$$Y = \beta S + \Gamma_1 X + \varepsilon_Y \tag{2.13}$$

$$S = \Gamma_2 X + \varepsilon_S. \tag{2.14}$$

The error terms ε_i, $i \in \{Y, S\}$, are assumed to be mean zero, mutually independent and independent of S, Y, and the elements of X; it is in this sense that these equations are structural. The presence of X in both Equations (2.13) and (2.14) captures the possibility for confounding via the path $S \leftarrow X \rightarrow Y$. I retain the structure of selection assumed in the main text, namely that latent utility C^* is a linear function of S, Y, and a mean-zero structural error term:

$$C^* = \alpha_S S + \alpha_Y Y + \varepsilon_{C^*}, \tag{2.15}$$

where $\alpha_i \in \mathbb{R}$, $i \in \{Y, S\}$. There is therefore no differential selection on other confounders X; an individual is assumed to return migrate, $C = 1$, if latent utility is below some fixed threshold:

$$C(S, Y) = \begin{cases} 1 & \text{if } C^* < K, \\ 0 & \text{otherwise.} \end{cases} \tag{2.16}$$

The assumption that the structural equations are linear implies that the true parameter β is proportional to the covariance of Y and the residualised version of S, given the covariates X:[15]

$$\beta = \mathrm{E}[\varepsilon_S^2]^{-1}\mathrm{E}[Y\varepsilon_S] \tag{2.17}$$

[15] To verify Equation (2.17), substitute Equation (2.13) into Equation (2.17) and note that X and ε_S are uncorrelated by assumption.

However, conditioning on no return migration, $C = 0$, when estimating Equation (2.13) means that the OLS estimand is instead:

$$\widehat{\beta} = \mathrm{E}[\varepsilon_S^2 | C = 0]^{-1} \mathrm{E}[Y \varepsilon_S | C = 0]. \tag{2.18}$$

Substituting Equation (2.13) into equation (2.19) and rearranging terms, one can show that

$$\begin{aligned} \widehat{\beta} &= \beta + \mathrm{E}[\varepsilon_S^2 | C = 0]^{-1} \mathrm{E}[((\beta \Gamma_2 + \Gamma_1) X + \varepsilon_Y) \varepsilon_S | C = 0] \\ &= \beta + \mathrm{E}[\varepsilon_S^2 | C = 0]^{-1} \mathrm{E}[(Y - \varepsilon_S) \varepsilon_S | C = 0]. \end{aligned} \tag{2.19}$$

The sign of the bias in $\widehat{\beta}$ relative to the true causal effect β is given by the term

$$\begin{aligned} \mathrm{E}[(Y - \varepsilon_S)\varepsilon_S | C = 0] &= \int y \mathrm{E}[\varepsilon_S | C = 0, y) dF_Y(y | C = 0) - E(\varepsilon_S^2 | C = 0) \\ &= \int y \mathrm{E}[\varepsilon_S | C* \geq K, Y = y) dF_Y(y | C^* \geq K) \\ &\quad - E(\varepsilon_S^2 | C^* \geq K) \\ &= \int y \mathrm{E}[\varepsilon_S | \alpha_S S + \alpha_Y y + \varepsilon_{C^*} \geq K) dF_Y(y | C^* \geq K) \\ &\quad - E(\varepsilon_S^2 | C^* \geq K) \\ &= \int y \mathrm{E}[\varepsilon_S | \alpha_S \varepsilon_S \geq K - \varepsilon_{C^*} - \alpha_Y y - \alpha_S \Gamma_2 X) dF_Y(y | C^* \geq K) \\ &\quad - E(\varepsilon_S^2 | C^* \geq K). \end{aligned} \tag{2.20}$$

The second term of (2.20) is the expectation of a positive random variable, it is therefore negative. The expectation under the integral in the first term is the expectation of a mean-zero random variable conditional on the distribution being truncated. If $\alpha_S < 0$, i.e. $\mathrm{Cov}(S, C) < 0$, then the distribution will be right truncated and the expectation will negative, implying, since $Y \geq 0$, that the integral will be positive and the total bias of $\widehat{\beta}$ relative to β is negative. If, on the other hand, $\alpha_S > 0$, then the distribution of ε_S is left-truncated. The expectation under the integral will be positive and the bias cannot, in general, be signed; it will be a function of the full joint distribution of the data.

2.A.3 Potential outcomes formulation

I derive an alternative expression for the bias induced by selective return migration using the potential outcomes framework. Consider a simplified set-up in which the immigrant's initial firm can be either low-conational share, $S = 0$, or high-conational share, $S = 1$. The outcome of interest is an immigrant's subsequent employment rate, $Y \in [0, 1]$, and whether they have left the country by the end of the sample period, which leads to truncation, $C = 1$, or not, $C = 0$. Both employment and return migration are a function of potential outcomes given S:

$$Y = SY^1 + (1 - S)Y^0 \tag{2.21}$$

$$C = SC^1 + (1 - S)C^0. \tag{2.22}$$

I am interested in the causal effect of starting out in a high-conational share firm on the subsequent employment rate, $\mathrm{E}[Y^1 - Y^0]$, and I only observe individuals who have not left the country at the end of the sample period, $C = 0$. To focus on the bias induced by selection on the treatment, S, into return migration, suppose that the tuple $\{Y^0, Y^1, C^0, C^1\}$ is independent of S, conditional on some set of observed controls, X. The marginal effect of the initial conational share on subsequent employment rates estimated in the regressions presented in Section 2.4.2.1 is a parametric estimate of the difference in employment rates between observed individuals who started in a high-conational share firm and observed individuals who started in a low-conational share firm, conditional on controls:

$$\mathrm{E}[Y|X, S = 1, C = 0] - \mathrm{E}[Y|X, S = 0, C = 0]. \tag{2.23}$$

Re-writing this expression as a function of potential outcomes we obtain

$$\begin{aligned}
&\mathrm{E}[Y|X,S=1,C=0]-\mathrm{E}[Y|X,S=0,C=0]\\
&\quad=\mathrm{E}[Y^1|X,S=1,C^1=0]-\mathrm{E}[Y^0|X,S=0,C^0=0]\\
&\quad=\mathrm{E}[Y^1|X,C^1=0]-\mathrm{E}[Y^0|X,C^0=0]\\
&\quad=\underbrace{\mathrm{E}[Y^1-Y^0|X,C^1=0]}_{\text{causal effect}}+\underbrace{\mathrm{E}[Y^0|X,C^1=0]-\mathrm{E}[Y^0|X,C^0=0]}_{\text{selection bias}},
\end{aligned} \tag{2.24}$$

where the first equality follows from the definitions of Y and C, the second from the conditionally random assignment of S, and the third is obtained by adding and subtracting $\mathrm{E}[Y^0|X,C^1=0]$ and using the linearity of the expectations operator. Equation (2.24) illustrates the nature of the identification problem created by selection into return migration. The observational difference can be broken into two terms. The first term, $\mathrm{E}[Y^1-Y^0|X,C^1=0]$, is a causal effect, though for a specific subpopulation: individuals who would not leave the country if they started out in a high-conational firm, $C^1=0$.[16]

The remaining terms of Equation (2.24) reflect selection into return migration caused by the initial conational share. If individuals who stay in the country when starting out in a high-conational share firm ($C^1=0$) would have had higher subsequent employment rates on average had they started out in a low-conational share firm (Y^0) than those individuals who stay in the country when they start out in a low-conational share firm ($C^0=0$), this term will be positive. This might be the case if starting in a firm with a high-conational share makes individuals with a weaker baseline employment potential more likely to leave the country, say because knowing fewer natives at the start makes it harder for them to integrate, learn German, navigate administrative procedures, find and

[16]In my parametric estimates in section 2.4.2 I further assume that the causal effect is constant over values of X, so the first term of equation (2.24) simplifies to $\mathrm{E}[Y^1-Y^0|C^1=0]$, an average treatment effect (ATE), rather than a conditional average treatment effect (CATE), i.e. an ATE conditional on X.

change accommodation, etc. On the other hand, if starting out in a high-conational share firm makes individuals with a low Y^0 more likely to stay in Germany, perhaps because they feel more at home in Germany since they don't have to speak German in the workplace, then the bias term will be negative.

It is also possible that the initial conational share has no effect on the decision to return home, and $C^1 = C^0 = C$.[17] In this case, the selection bias term is zero and the causal effect can be estimated as the difference in outcomes for observed individuals. Furthermore, the causal effect estimated is now $\mathrm{E}[Y^1 - Y^0|X, C = 0]$, the causal effect of the initial conational share on earnings for all stayers, and not only individuals who stay when they start out in a high-conational firm ($C^1 = 0$).

Equation (2.24) sets out the nature of the identification problem created by selective return migration. In particular, it clarifies that this identification problem is conceptually independent from any potential selection into initial conational share based on the controls X, i.e. a failure of the assumption that $\{Y^0, Y^1, C^0, C^1\}$ is independent of S conditional on X. However, it is difficult to think through the potential sign of the bias created, much less evaluate it empirically, since it depends on a fundamentally unobservable, counterfactual, quantity: the employment rate that individuals who do not return home when starting in a *high*-conational share firm would have had, had they started out in a *low*-conational share firm, $Y^0|C^1 = 0$.

[17] This assumes there is no individual-level effect of starting conational share on the subsequent return migration decision. The absence of causal effect is often taken to mean that there is no effect on average, $\mathrm{E}[C^1 - C^0] = 0$. The stronger formulation here simplifies the exposition.

3

THE LABOUR MARKET OUTCOMES AND ASSIMILATION OF URBAN MIGRANTS IN INDONESIA

3.1 INTRODUCTION

Large and persistent gaps in standards of living have been widely documented between rural and urban locations in developing countries. These gaps are present not only in wages and consumption, but also in access to running water and electricity, nutrition, or mortality (Lagakos, 2020). In spite of these gaps, in many regions of the developing world, including sub-Saharan Africa, South Asia, or Southeast Asia, less than half the population resides in urban areas. The failure of more individuals to take advantage of the opportunity to arbitrage away rural-urban wage differences by moving has been labelled a puzzle (Gollin et al., 2014; Henderson and Turner, 2020).

However, the presence of large wage gaps between urban and rural locations is not by itself evidence that individuals would experience earnings gains if they moved. In frictional labour markets, prospective migrants also need to form beliefs about the probability that they will find a job if they move, the quality of job available to them in the urban labour market, and their chances of progressing up the job ladder before deciding whether or not to move. In this chapter I study the labour market outcomes of rural-urban migrants in Indonesia, the world's fourth-most populous country; I set out to provide empirical evidence on whether moving helps or harms migrants in the labour market.

Indonesia is in the middle of the process of urbanisation; 56 per cent of its population resided in an urban area in 2018, making it more urbanised than most countries in sub-Saharan Africa or south Asia, but less urbanised than most of Latin America or the Middle East. I focus on the period from 1988, when Indonesia's urbanisation rate was 29 per cent, to 2015, when its urbanisation rate was 53 per cent. I will focus in particular on two aspects of the labour market outcomes of migrants: the ease of finding and staying in a job, and success or otherwise in moving up the job ladder.

Studies of rural-urban migration in the developing world typically consider the decision to migrate as being made by comparing potentially idiosyncratic, but static, differences in wages between rural and urban locations (Pulido and Święcki, 2019; Bryan and Morten, 2019; Lagakos et al., 2020). On the basis of these comparisons, they conclude that migration costs are an important part of the explanation of the persistence of rural-urban wage gaps.

In principle, however, urban migrants may have a harder time finding work if they move than if they stay. If migrants start out in the city with a period of unemployment, there may be an opportunity cost to moving, increasing the total migration cost. Research on temporary migration in Bangladesh has highlighted that migrants are not certain to find work (Bryan et al., 2014), while studies of international migration have also shown that international migrants are much less likely to be employed than natives (e.g. Sarvimäki, 2011). Conversely, if migrants have an easier time finding work in the city, this might indicate that migration costs are larger than simple comparisons of wages in both locations would imply. The first contribution of this chapter is to provide evidence that immigrants receive a transient employment boost at migration, suggesting that the migration cost may be larger than suggested by comparisons of wages alone.

In addition to potential differences in employment rates between urban and rural locations, the returns to migrating to larger cities are generally thought to be dynamic (De La Roca and Puga, 2017), at least in developed countries. The attraction of cities might be that they offer better job ladders, even if the entry position might be similar in both rural and urban locations. Static comparisons of earnings across locations, however, will miss any job ladder effects of moving. The second contribution of this chapter is to study the occupational dynamics of urban migrants. Migrants appear to enjoy more rapid occupational upgrading, suggesting that static comparisons of wages of short-term migrants and non-migrants might underestimate the longer-term returns to migrating. This would also lead researchers to underestimate the size of migration costs necessary to rationalise observed patterns of migration.

Finally, part of the earnings difference between international migrants and natives lies in the fact that human capital acquired in the home country earns a discounted return abroad (Eckstein and Weiss, 2010). Static comparisons of earnings across locations that assume that rural human capital is fully portable may overestimate the true return to migrating, also leading the researcher to conclude that migration costs are higher than they are. The third contribution of this chapter is, therefore, to show that urban migrants do not appear to suffer an employment or occupational penalty relative to urban natives. This finding provides support for the assumption that the returns to human capital are the same across locations.

Estimating the labour market performance of urban migrants is complicated by the fact that urban migrants are likely to be selected relative to the rural population as a whole. Furthermore, the decision to migrate is endogenous, and individuals likely migrate in response to localised shocks to sector-specific productivity or migration costs, making it difficult to identify the true return to migration. To address this identification problem, in the first part of the chapter I use a research design that compares urban migrants with siblings who did not migrate.

Comparing migrants with their siblings has several advantages. First, siblings share many unobserved or hard-to-observe characteristics, whether these are genetic or have to do with the environment in which they were raised, making siblings who didn't move an attractive counterfactual for urban migrants, had they not moved. Second, by focusing on siblings who were likely to reside together before the move, localised, time-varying productivity shocks or shocks to migration costs are common to movers and stayers, removing many possible time-varying sources of endogeneity. For these reasons, sibling comparisons have already been used to evaluate historical returns to international migration (Abramitzky et al., 2012). This work, in turn, builds on a long tradition in labour economics of comparing siblings to evaluate things such as the return to schooling (see Card, 1999, for an early review), or the effect of maternal behaviour on the outcomes of children (see e.g. Almond and Currie, 2011; Oster, 2019). Interpreting the return to migration relative to siblings is, however, complicated by the fact that the decision to send a migrant is often made at the household level (Gröger and Zylberberg, 2016; Munshi and Rosenzweig, 2016; Stark and Bloom, 1985), and, conditional on sending a migrant, there may be intra-household selection of migrants (Dustmann et al., 2017).

Using siblings as a control group, I conduct an event study in urban migrants' employment rates and occupational rank. I find that urban migrants benefit from migration, their employment rates being at least 10 percentage points higher immediately after migration than their siblings who stayed back. They are also employed in higher-ranked occupations. However, the employment gains are temporary, while the occupational gains are more persistent and, if anything, increase over time. While selective return migration complicates the assessment of the true long-term outcomes for all migrants, it does appear that migrants outperform their siblings in the longer run.

Comparing urban migrants to their siblings can tell us what the effect of urban migration is on a migrant's employment rate or occupation. However, to know how informative static earnings gaps between urban and rural locations are about the potential gains from urban migration, we need to understand whether urban migrants have similar outcomes to comparable urban natives when they do migrate. It might be the case that urban migrants, whose social networks are less developed in the urban area than urban natives, or whose prior work experience might be in occupations in short demand in the urban location, fare worse than comparable natives when they do migrate. To understand whether this is the case, in the second part of the chapter I use a matching strategy to identify an appropriate reference group of urban natives for urban migrants. As was the case when comparing them with siblings, urban migrants are more likely to be employed than urban natives immediately after arrival. The employment gains, however, are short-lived. There is also little evidence that urban migrants either outperform or under-perform relative to natives in occupational rank in the longer term. Taken together, the results of this chapter suggest that urban migrants, unlike international migrants, do not face a labour market penalty when migrating to urban areas.

The chapter proceeds as follow. In Section 3.2 I briefly review the main relevant literature on rural-urban migration. In Section 3.3 I present the data used in this project. In Section 3.4 I present my event study, which compares urban migrants with siblings who stayed back, to estimate the labour market return to migrating. In Section 3.5 I present my matching approach to estimating differences in labour market outcomes between urban migrants and natives. Section 3.6 concludes and discusses directions for future research.

3.2 LITERATURE

Here I selectively review the main relevant papers on rural-urban migration in developing countries. Previous explanations of the relative lack of rural-urban migration in developing countries have focused broadly on the relative importance of sorting on skill between rural and urban locations in explaining observed wage differences (Lagakos and Waugh, 2013; Young, 2013) versus barriers to mobility (Bryan et al., 2014; Munshi and Rosenzweig, 2016). On balance, selection on skills cannot fully account for observed internal migration patterns, which require nontrivial migration costs to explain observed patterns of immobility (Bryan and Morten, 2019; Pulido and Święcki, 2019), or indeed a combination of migration costs and a bias towards living where one is born (Zerecero, 2020). Hicks et al. (2017) provide a dissenting view, arguing that the returns to urban migration in Indonesia and Kenya are negligible, and entirely due to individual selection, explaining why more individuals do not move. On the other hand, Sarvimäki et al. (2020) show that forced migrants in Finland who moved from agriculture to urban areas were better off after the move. However, Lagakos et al. (2020) show how heterogeneity in the costs and benefits of migration mean that the return to migration cannot be inferred from observational returns without knowing the joint distribution of costs and benefits, complicating the interpretation of the empirical evidence.

Classic studies of migration treated the decision to migrate as an individual one made on the basis of relative wages in rural and urban locations (Harris and Todaro, 1970). However, since at least the work of Stark and Bloom (1985), many studies have shown that the decision to send a migrant is often taken at the household level to mitigate against the risk of income shocks, whether ex-ante (Rosenzweig and Stark, 1989; Munshi and Rosenzweig, 2016) or ex-post (Gröger and Zylberberg, 2016; Gröger, 2021). In light of the insurance role of migration, these studies highlight that there is selection across households into migration. There is also selection within households (Dustmann et al., 2017), which will be important to consider when comparing siblings.

3.3 DATA

The data for this project come from the Indonesian Family Life Survey (IFLS). The IFLS is a panel survey that is notable for placing a particular emphasis on tracking down adult respondents across waves even in the event that a respondent leaves the household and changes address. The first wave of the IFLS was conducted in 1993-94; it included 33,081 individuals in 7,244 households from 13 Indonesian provinces who were representative of the 83 per cent of the population of Indonesia in 1993. Subsequent waves were carried out in 1997, 2000, 2007-08, and 2014-15. Recontact rates for adult respondents from the first wave were 82 per cent in the final wave, including individuals who had died in between; the final wave included 58,325 individuals in 16,931 households, and 83,700 individuals are covered by at least one wave of the survey. A full overview of the IFLS is available in Strauss et al. (2016).

Certain individuals in IFLS households are targeted for detailed interviews on their contemporaneous employment situation and a shorter interview on their employment and migration history since the previous wave.[1] Table 3.1 reports summary statistics for all individuals in the IFLS data. Statistics Indonesia (BPS) classifies each village (*desa*/*kelurahan*) as rural or urban; I report summary statistics by urban status. In Panel A, I report differences in employment rates between rural and urban locations, and the total number of observations (individual by survey wave). Employment rates are moderately higher in rural areas, by around three percentage points, and the number of respondents is fairly evenly balanced between rural and urban locations. In Panel B I

[1]The precise criteria to be eligible for a detailed interview vary from wave to wave, but are designed to cover most individuals aged 15 and over in a household, and all individuals having previously provided detailed information.

show that the occupational distribution, conditional on employment, is quite different over rural and urban areas. Over half of rural respondents are employed in agriculture and related activities, compared with ten per cent in urban areas, while urban residents are much more likely to be employed in clerical and administrative activities, sales, and low-skilled services such as housekeeping, maids, or hospitality. Panel C reports the time-invariant highest level of schooling completed. The education distribution of urban residents first order stochastically dominates that of rural residents.

These static differences between rural and urban residents already suggest that rural residents might not migrate more simply because they have a higher chance of being employed in rural areas. This descriptive evidence, however, is not conclusive; in particular it does not tell us whether a given urban migrant faces a higher probability of being unemployed than they would have had they stayed back. Furthermore, given that rural residents have less education and experience in different occupations, this evidence suggests that rural migrants might lack the human capital necessary to take up urban jobs, which might further lower their employment rates when they migrate. Again, however, this evidence is only descriptive, and does not tell us whether urban migrants are less likely to be employed than urban natives with similar human capital, and if they are employed in different occupations.

To answer these questions, I will draw on the employment and migration histories in the IFLS. A subset of respondents are asked to fill out year-by-year information, back to the year of the previous survey wave, on any migration spells where an individual crosses village borders to live for at least six months, ruling out seasonal migration, and their primary activity during the year. The retrospective information is less detailed than the information on an individual's location or employment in the survey year. The migration history does not include the BPS classification of an individual's location, only an individual's self-reported classification of their then-location of residence as a village, town, or

Table 3.1: Summary of rural-urban differences

	(1) Urban	(2) Rural	(3) Difference
Panel A			
Unemployed	0.501	0.468	-0.033
Employed	0.499	0.532	0.033
Observations	88,132	83,133	171,275
Panel B			
Technical	0.021	0.007	-0.014
Prof.	0.059	0.035	-0.024
Administrative/Mangerial	0.007	0.002	-0.005
Clerical	0.071	0.019	-0.053
Sales	0.242	0.134	-0.108
Low-skill S.	0.188	0.073	-0.115
Agric.	0.106	0.538	0.433
Unskilled P.	0.101	0.063	-0.039
Skilled P.	0.044	0.021	-0.023
Observations	41,864	41,807	83,673
Panel C			
No formal schooling	0.349	0.436	0.086
Elementary school	0.183	0.249	0.066
Junior high school	0.148	0.159	0.011
Senior high school	0.199	0.106	-0.093
College (D1,D2,D3)	0.064	0.027	-0.036
University (BA, MA, PhD)	0.057	0.024	-0.033
Observations	46,494	37,213	83,751

Note: Summary statistics for all IFLS respondents by rural-urban status. Information drawn from contemporaneous modules of the IFLS, waves 1-5. All differences are significant at the 1 per cent level.

city. I will classify a location as rural if the respondent describes it as a village, and urban if it is a town or city. Respondents are only asked for retrospective earnings information in the first three waves. I will therefore focus on the employment and occupational choice margins, for which retrospective information is available in all waves.

In Table 3.2 I report summary information on the number of rural-urban moves an individual reports making during the sample period in column one, and the duration of each spell in an urban location in column two. I restrict the sample to individuals who report being born in a village, residing in a village at age twelve, and who first appear in a village in my dataset, and censor both the number of moves and the duration of a move at five. 80 per cent of rural residents never move to an urban location and only 8 per cent move permanently to an urban location (though they may move between urban locations). A further nine percent move to an urban location and then move back to a rural location, though not necessarily the one where they started, while the remaining individuals move multiple times between rural and urban locations. The average move is also relatively short; 39 per cent of movers stay two years or less, and only 33 per cent stay more than five years. These data suggest that the distinction between temporary or seasonal migration and permanent migration may not be so relevant in this context, with migration spells existing on a continuum.

3.4 THE LABOUR MARKET PERFORMANCE OF URBAN MIGRANTS

Most rural residents never migrate to the city. Those who do are likely to be different from other rural residents, suggesting that straightforward comparisons of movers and stayers may be biased, likely overestimating the effect of moving on an individual's labour market outcomes. To overcome this difficulty, I compare movers with their siblings who stayed

Table 3.2: Summary urban migration rates

	# Moves	Duration
0	79.64	
1	8.05	19.61
2	9.20	19.44
3	1.68	16.15
4	1.12	11.99
5	0.32	32.81
N	26662	6105

Note: Information drawn from the employment and migration history modules of the IFLS, waves 1-5. Sample is restricted to individuals who self-report being born and raised in a village, and are living in a village when first interviewed.

back, using an event study design.

3.4.1 *Sample selection*

Groups of siblings are identified from the household roster. In each survey wave I define three partitions of the full set of individual observations: (i) individuals listed as children of the same household head; (ii) individuals listed as the household head or head's siblings; and (iii) individuals listed as the household head's spouse or the head's siblings-in-law. The partition of all respondents into groups of siblings is then defined as the finest common coarsening of these three partitions, across all survey waves.

To focus on the effect of migration as an adult for rural residents, I restrict my sample to individuals who report they were (i) born in a village; (ii) lived in a village at age 12; and (iii) are in a village when first observed in my sample. I retain all observations where individuals are aged 12-60. I then limit attention to groups of siblings that contain between two and eight individuals satisfying the aforementioned restrictions and at least one non-mover. I will separately analyse movers who are either the first or the only one of their siblings to move and movers who move after another sibling does so. Furthermore, movers and stayers must be observed in both the year before the move and the year of the move. These restrictions leave me with a full sample of 3,033 individuals: 1,419 movers and 1,614 stayers.

The empirical design is an event study; therefore, let t refer to calendar time, d refer to the event year, and k refer to event time, i.e. $k = t - d$. The event of interest is defined as an individual's first move to an urban location. 53 per cent of my sample never move, 15 per cent move once, 25 per cent twice (i.e. to an urban location and back), and 7 per cent more than twice. Movers can therefore broadly be separated into those who move and stay, and those who move once and then move back to a rural location, which may not necessarily be the rural location of origin. Since urban migration may have labour market effects that last beyond a return to the village, migrants are considered migrants regardless of whether they are still at the urban location or whether they have moved back. Some stayers have multiple siblings who emigrate. These stayers serve as control observations for each event, which means that a stayer may appear multiple times in my dataset in a given calendar year. The total number of non-unique stayers is therefore 2,126.

Table 3.3 contains summary statistics for movers and stayers in the year before the emigration, i.e. $k = -1$. The identifying assumption for an event study is that the outcome of interest follows a parallel trend for movers and stayers. Level differences in observed characteristics therefore do not invalidate the identifying assumption, they simply provide information about what type of individual moves. Movers and stayers are equally likely to be male as female, stayers are on average two years older than movers, and consequently somewhat lower in the birth order. Stayers score 0.1 of a standard deviation lower on the Raven test of cognitive skills, a significant difference, and they tend to be somewhat less educated. Note, however, that there is a clear relationship between age at which the test was taken and Raven test scores in my data, which increase linearly from age 12 to around 16 and decline linearly thereafter. The Raven score does not significantly predict mover status (p-value = 0.48) after controlling for a set of dummies in age at which the test was taken. Stayers clearly have a higher employment rate than movers, although, conditional on employment, stayers have a slightly lower occupational rank on average.

Individuals are also asked for their reason for migrating. For both genders the most common reason for moving is work, for 44 per cent of women who move and 62 per cent of men. Unfortunately, the categories do not allow us to distinguish between individuals who are unemployed and move to find work, individuals who are employed and move to find better work, or individuals who move because they have already been offered work at the destination, whether they were employed or not beforehand. However, we may note that 59 per cent of men moving for work reasons are employed in $k = -1$, while only 32 per cent of women moving for the same reason are employed. While not conclusive proof, these numbers suggest that men may be more likely to move to find better work, while women may be more likely to move to find any work, though all types of work moves considered above are likely to be present to some degree in the data. Education is the next most common category for both genders, at 18 per cent for both men and women. Fewer than one per cent of movers report natural or other disasters, or sickness or death of other household members as the reason for moving, so these migrants can be considered representative of migrants in normal times (c.f. Gröger and Zylberberg, 2016, on the relatively worse performance of migrants who move in response to a natural disaster).

Table 3.3: Movers and sibling stayers in $d-1$

	(1) Stayers	(2) Movers	(3) Difference
P(Female)	0.51	0.50	-0.00
	(0.50)	(0.50)	(0.02)
Age	22.88	20.88	-1.99**
	(7.60)	(6.01)	(0.24)
Birth order	2.80	2.93	0.13*
	(1.71)	(1.66)	(0.06)
Cognitive score	0.08	0.17	0.10**
	(0.99)	(0.96)	(0.03)
P(Employed)	0.55	0.44	-0.11**
	(0.50)	(0.50)	(0.02)
Rank	20.98	22.78	1.80+
	(19.95)	(19.80)	(1.00)
No formal schooling	0.11	0.05	-0.05**
	(0.31)	(0.23)	(0.01)
Elementary school	0.26	0.22	-0.04**
	(0.44)	(0.41)	(0.01)
Junior high school	0.30	0.25	-0.05**
	(0.46)	(0.43)	(0.02)
Senior high school	0.24	0.31	0.07**
	(0.43)	(0.46)	(0.02)
College (D1,D2,D3)	0.05	0.09	0.04**
	(0.22)	(0.28)	(0.01)
University (BA, MA, PhD)	0.04	0.08	0.04**
	(0.20)	(0.27)	(0.01)
Observations	2,126	1,419	3,545

Note: Information drawn from the employment and migration history modules of the IFLS, waves 1-5. Compares individuals born and raised in a village who move with siblings who stay behind in event year $k = -1$. Stayers are not unique observations, an individual may be a stayer for multiple sibling movers, typically in different years. + p<0.1, * p<0.05, ** p<0.01.

3.4.2 *Empirical approach*

To study the effect of urban migration on labour market performance, I compare migrants with siblings who stayed behind in an event study. The main empirical specification is therefore as follows:

$$y_{ik} = \alpha_i + \alpha_k + \sum_{s=-5}^{10} \beta_s \times \mathbf{1}(mover_i = 1) \times \mathbf{1}(k = s) + \alpha_{t(ki)} + \alpha_{age_{ki}} + \varepsilon_{ik}. \tag{3.1}$$

The specification includes individual fixed effects α_i that control for level differences between movers and stayers, event time fixed effects α_k and calendar time fixed effects $\alpha_{t(ki)}$ and age (in years) fixed effects $\alpha_{age_{ki}}$. The parallel trends assumption implies that $\beta_k = 0$ for $k < 0$, while the return to migrating is measured by β_k for $k \geq 0$. Furthermore, the coefficients β_k are normalised relative to β_{-2}.[2] The main outcomes I focus on are an indicator variable for being employed in event year k, and the occupational rank in year k, where unemployment has the lowest rank, 0. Standard errors are clustered by sibling group level.[3]

This event study improves on prior empirical work estimating rural-urban migration using the IFLS data in several ways. First, two papers estimate static productivity wage differences, conditional on employment, between rural and urban locations (Lagakos et al., 2020; Pulido and Święcki, 2019). However, in developed countries, moving from a smaller to a larger city is thought to lead to both a one-time jump in earnings and to higher earnings growth (De La Roca and Puga, 2017); a dynamic specification is needed to test whether this is the case in Indonesia.

Second, Hicks et al. (2017) include a dynamic specification with individual fixed effects, similar to the the one presented in Equation

[2]See Borusyak and Jaravel (2017) on identification problems in event studies, and in particular the difficulty of separately identifying level effects and trend effects.

[3]Note a stayer may appear multiple times in the dataset as a control observation for different sibling movers. For the purposes of calculating individual fixed effects, repeated observations are treated as separate observations, but will belong to the same sibling cluster for calculating standard errors.

(3.1). However, they do not define a control group of non-movers, the effect of rural-urban migration is identified using movers who have not yet moved as a control group. This type of empirical specification will be correct if there is only a level effect of moving, but will be biased if moving also has an effect on the trend in the outcome y_{ik} (Azoulay et al., 2010). Furthermore, Hicks et al. (2017) focus on earnings conditional on being employed, a specification that suffers from selection bias, since urban migration, even if randomly assigned, changes the composition of the group of employed individuals (see e.g. Angrist and Pischke, 2009). If individuals who are employed when in an urban area would have had lower earnings, on average, in a rural area than individuals who are employed when in a rural area, the selection bias is negative, potentially explaining the null effect estimated by Hicks et al. (2017). Such a pattern of selection would be consistent with sorting based on comparative advantage in rural or urban work.

Third, both the static and dynamic fixed effect specifications discussed above are potentially biased if the timing of migration coincides with time-varying local productivity shocks that affect both the destination and the origin. For example, infrastructure improvements in a district might induce an individual to migrate to the nearest town where they earn higher wages, but they may also improve economic outcomes at the rural origin location. Conversely, negative local productivity shocks, such as bad seasonal rainfall, could simultaneously induce some individuals to migrate and depress employment rates for stayers. Alternatively, household-level shocks such as the death or illness of a main earner might simultaneously induce increased employment rates for all working-age household members and induce some (e.g. younger) household members to move to an urban area to look for work. These types of shocks are a relevant concern when the control group is movers who have not yet moved, who typically come from different locations and household to movers in a given year. In my specification, the control group are siblings who never move. These siblings typically co-resided with the mover before the migration event, so any effect I find is robust to time-varying origin household-level productivity or employment shocks.

The main identifying assumption for this empirical design will be that outcomes between movers and non-movers would have followed parallel trends, had the mover stayed back. While it has been observed that there is within-household selection of migrants based on fixed characteristics, such as risk aversion (Dustmann et al., 2017), these in principal do not threaten identification in this case. Instead, threats to identification will come from time-varying individual-level shocks to either the cost or benefit of migration. Such shocks could arise if, for example, individuals migrate in response to individual-specific shocks to future employment prospects in rural areas. The estimated effect could also be biased if there are spillovers from mover siblings to stayer siblings. For example, remittances from the mover back to the origin household might reduce the incentive for stayers to work, although the income levels of the household considered make this unlikely. Alternatively, moving might be the result of household-level bargaining, where the mover goes to find work in the city and the stayers are expected to care for old parents or young children, lowering labour-force participation. Formally, the Stable Unit Treatment Value Assumption (SUTVA) may fail to hold. I will show evidence that there does not appear to be a trend break in the outcomes of stayers when a sibling migrates.

Sample selection issues also need to be taken into account when evaluating the results of this approach. Many papers have noted that there is selection across households into migration (e.g Rosenzweig and Stark, 1989; Munshi and Rosenzweig, 2016; Gröger and Zylberberg, 2016). Since only a small proportion of rural households send a migrant, the estimated effect of migrating can only be interpreted as an effect of treatment on the treated and may not be representative of the effect for a migrant drawn from the typical rural household. Even among migrants, the sample studied may be positively selected. Indeed the IFLS asks respondents about migration episodes lasting more than six months; it is likely that individuals who fail to find work in the city return home before six months have passed, leading me to exclude them from my sample. The nature of the data do not allow me to assess how serious this selection problem may be. Finally, my event study only identifies a partial equilibrium effect, it is silent on the hypothetical effect of urban migration in the context of large aggregate population movements.

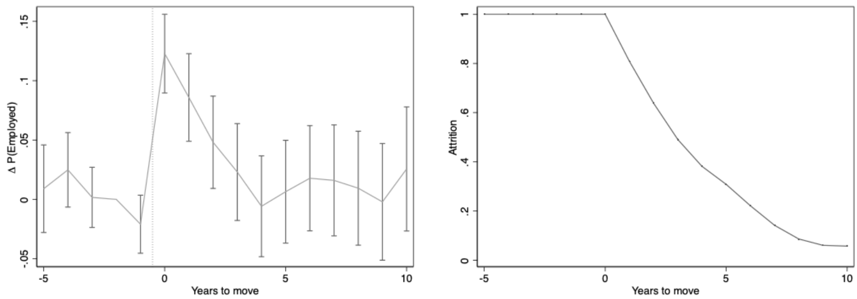
I focus on a 15-year window around the event year, $k \in [-5; 10]$. The panel is unbalanced as some individuals are not observed in every year for various reasons. They may be too young to be included in the full pre-event period, they may have moved less than 10 years before the last survey wave was conducted, or they may not have been located for a follow-up interview. The latter case in particular could lead to selection bias, if less successful movers move on again and are not found. For this reason, I will also report results for a subset of the data for which a shorter, balanced panel is available for movers and stayers for $k \in [-2; 6]$.

3.4.3 *Employment results*

3.4.3.1 Main results

The first measure of labour market performance I consider is an individual's employment rate, measured as a dummy variable for reporting being employed in event year k. The left panel of Figure 3.1 reports the β_s coefficients defined in Equation (3.1), which measure the difference in employment rates between movers and stayers, relative to the difference in the base year, $k = -2$, and 95 per cent confidence intervals. Several patterns are worth noting. First, there is evidence of a modest Ashenfelter dip before an individual moves to the urban area. Movers are 2.1 percentage points less likely to be employed than stayers in period $k = -1$ relative to period $k = -2$, although the differences in employment rates between both groups are statistically indistinguishable from one another in the three years preceding the move. Second, urban migration is associated with large initial gains in employment. Movers' employment rates are 12.3 percentage points higher, relative to stayers, in the year of the move than two years before the move, when the employment rate for stayers is on average 53.8 per cent. Third, the effect appears to be transient; three years after the move differences in employment rates between movers and stayers are no longer significantly different from the difference in $k = -2$ at the five per cent level.

Figure 3.1: Employment effect of urban migration



Notes: Left panel: effect of moving to an urban location on employment rates. The effect is normalised relative to β_{-2}, 95 per cent confidence intervals are shown. Right panel: attrition from urban location, i.e. return migration among urban migrants. See text for sample definition.

The values of the coefficients and standard errors are also reported in column 1 of Table 3.4. To assess whether there is selection into migration within groups of siblings, column 2 of Table 3.4 reports results of a less stringent specification, which includes sibling fixed effects, rather than individual fixed effects, as in Abramitzky et al. (2012). If mover and stayer siblings have different unobserved time-invariant characteristics that are systematically related to their employment outcomes, the coefficients from this specification will be biased, while the coefficients from the specification with individual fixed effects, reported in Figure 3.1 and again in column 1 of Table 3.4 will not be. The estimated coefficients for the specification with sibling fixed effects are lower than those when using individual fixed effects, particularly for $k < 5$. This confirms that, as in other settings (Dustmann et al., 2017), there is selection on time-invariant characteristics within households into migration.

While the declining time profile of effects of urban migration show that there is no permanent employment benefit to urban migration, it is possible that there is a constant employment benefit while still in the city that is obscured by return migration. Indeed, the right panel of Figure 3.1 shows that attrition from the urban location is high, mirroring the declining time profile of urban location employment effects. In column 3 of Table 3.4 I therefore repeat my analysis, conditioning on individuals still being in the urban location. Urban employment outcomes will shape the decision to return migrate. The employment effect of urban migration when conditioning on not returning will therefore be subject to selection bias analogous to the use of a bad control, so the estimates cannot be given a causal interpretation. Nevertheless, it is instructive to note that the same pattern of declining effects can be observed; the estimates of β_s are typically within one percentage point of the estimates for the full sample. The only exception is when $k \geq 9$, however the sample is quite small at this horizon and the effect is not significant. Given that one would expect return migrants to be negatively selected with respect to employment outcomes in the city, the absence of a negative employment effect for stayers suggests there is no long term employment effect of moving to the city.

An analogous source of selection bias will exist if either more or less successful individuals selectively move to new locations where they can't be tracked in a follow-up wave of the survey. There will then be selective truncation of the sample. This form of selection is also endogenous, so estimates of the employment effect of urban migration conditioning on no truncation likewise cannot be causally interpreted. Nevertheless, in column 4 of Table 3.4 I report estimates for a balanced panel on a smaller window, for $-2 \leq k \leq 6$. There is less evidence of an Ashenfelter dip for this shortened window and the estimated immediate effect is

Table 3.4: Employment rates of movers relative to sibling stayers

	(1) Individ. FE	(2) Sibling FE	(3) No return	(4) Balanced	(5) No educ.
Movers × k = -5	0.009	-0.038*	0.012		0.006
	(0.019)	(0.018)	(0.019)		(0.021)
Movers × k = -4	0.025	-0.019	0.028+		0.021
	(0.016)	(0.017)	(0.016)		(0.018)
Movers × k = -3	0.002	-0.043**	0.004		-0.006
	(0.013)	(0.016)	(0.013)		(0.015)
Movers × k = -1	-0.021+	-0.061**	-0.027*	-0.011	-0.004
	(0.012)	(0.016)	(0.012)	(0.016)	(0.014)
Movers × k = 0	0.123**	0.083**	0.116**	0.106**	0.183**
	(0.017)	(0.016)	(0.017)	(0.022)	(0.019)
Movers × k = 1	0.086**	0.046**	0.086**	0.071**	0.160**
	(0.019)	(0.017)	(0.020)	(0.026)	(0.021)
Movers × k = 2	0.048*	0.011	0.032	0.036	0.116**
	(0.020)	(0.017)	(0.022)	(0.027)	(0.022)
Movers × k = 3	0.023	-0.010	0.024	0.028	0.078**
	(0.021)	(0.018)	(0.024)	(0.027)	(0.023)
Movers × k = 4	-0.006	-0.034+	-0.020	0.004	0.023
	(0.023)	(0.019)	(0.028)	(0.028)	(0.024)
Movers × k = 5	0.006	-0.026	-0.004	0.021	0.032
	(0.022)	(0.019)	(0.029)	(0.027)	(0.024)
Movers × k = 6	0.018	-0.008	0.022	0.037	0.036
	(0.023)	(0.020)	(0.033)	(0.027)	(0.025)
Movers × k = 7	0.016	-0.010	-0.016		0.017
	(0.025)	(0.022)	(0.043)		(0.027)
Movers × k = 8	0.009	-0.017	-0.015		0.008
	(0.026)	(0.023)	(0.054)		(0.029)
Movers × k = 9	-0.002	-0.029	0.040		-0.010
	(0.026)	(0.023)	(0.070)		(0.029)
Movers × k = 10	0.026	-0.005	0.073		0.003
	(0.028)	(0.025)	(0.079)		(0.031)
N	41220	41220	35642	16419	34131
Clusters	1091	1091	1091	744	930
R^2	0.56	0.37	0.59	0.57	0.55

Note: Column one uses individual fixed effects, column two uses sibling fixed effects, column 3 restricts attention to individuals who have not return migrated, column 4 restricts attention to a balanced panel, column 5 excludes individuals and their siblings who report migrating for educational reasons. Coefficients are normalised relative to β_{-2}, which is omitted, standard errors are clustered by sibling group. + p<0.1, * p<0.05, ** p<0.01.

slightly smaller, at 10.6 percentage points, for this subsample, however the medium-term effects appear no smaller, suggesting that selective sample attrition is not a major concern.

In column 5 of Table 3.4, I confirm that the estimated effect does not simply reflect individuals who move for educational reasons subsequently becoming more likely to be employed as a result of education, instead of urban migration per se. I remove individuals who moved for education and their siblings and re-estimate my main specification. The estimated short-term effects are somewhat larger than in the main specification, consistent with individuals in education not working, but are still indistinguishable from zero in the longer-term.

Finally, investigating the common event time fixed effects, α_k (not reported here), fails to provide any evidence of a break or trend in employment that is common to both groups. The estimated fixed effects are indistinguishable from zero at the one per cent level. Furthermore, in results available on request, average employment for stayers is if anything increasing during the post-treatment period, both unconditionally, and conditional on age and year effects, failing to provide evidence of any negative spillover from treated to untreated units that could positively bias the estimated effect.

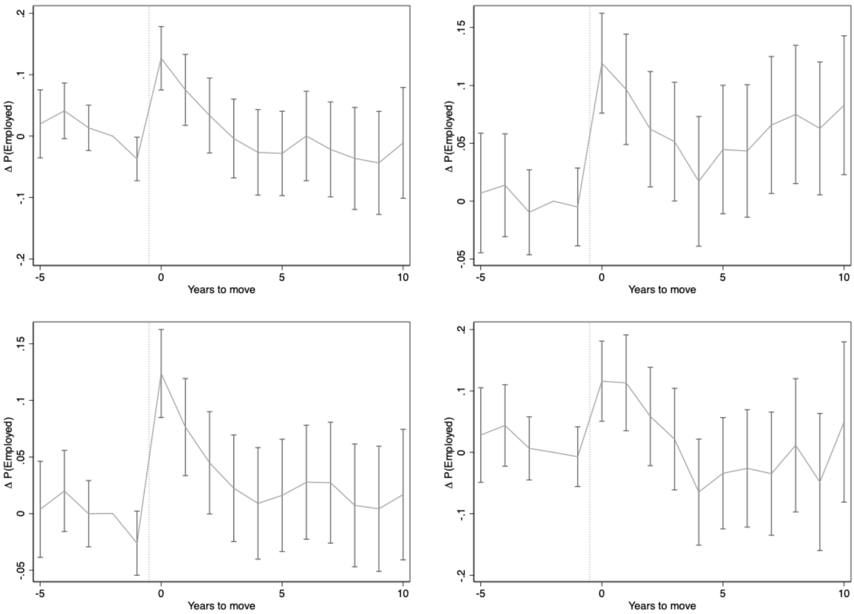
It might appear obvious that urban migrants experience, on average, a boost to their employment rates upon migration, given that many presumably migrate with the expectation that they are reasonably likely to find a job at the destination. Furthermore, if there are migration costs (Pulido and Święcki, 2019; Lagakos et al., 2020) or a birth-place bias (Zerecero, 2020) then movers must expect some return to moving, which might come in the form of temporarily higher employment rates. However, there is no evidence that *international* migrants, many of whom also migrate for economic reasons and arguably face higher costs, experience similar boosts to their employment rates, if anything the opposite may be true (Sarvimäki, 2011; Ansala et al., 2021). Furthermore, Farré and Fasani (2012) have shown that internal migrants in Indonesia typically have over-optimistic expectations about the return to migrating, although the return could still be relatively high, and I have documented that employment rates are, on average, higher in rural locations in the IFLS data. Theoretically, the fact that migrants may lack social networks at the destination and, in a linguistically diverse country such as Indonesia, may not be fluent in the local language at destination also suggest that they may face difficulties in finding a job at the destination. While it is true that the finding of positive earnings effects may be partly due to sample selection issues, if unsuccessful migrants stay less than six months, so don't report migrating, the finding that employment rates largely increase with migration is nevertheless economically significant.

3.4.3.2 Heterogeneous effects

I have already noted that women moving for work were more likely to be unemployed than men who stated they were moving for work. In the top panel of Figure 3.2 I separately estimate the effect for women (left panel) and men (right panel). The same estimates are also in columns 1 and 2 of Table 3.A.1. The pattern for women is similar to the overall pattern—there is evidence of a pre-trend, and the positive employment effect is not significantly different from zero for $k \geq 3$. For men, however, there is little evidence of a pre-trend, and there are persistent long-term effects. Women are also 5.8 percentage points less likely to re-migrate than men after three years, although the difference is not persistent and disappears after six years.

The differing patterns might be explained by the kind of occupations in which women tend to find employment in urban areas. For both genders, the most common two occupations are agricultural and animal husbandry workers followed by salesmen/women and shop assistants. For women, however, the next three occupations are maids and housekeepers, tailors and dressmakers, and working proprietors of catering and lodging services, while for men the top five is rounded out by bricklayers and

Figure 3.2: Heterogeneous employment effects of urban migration



Notes: Heterogeneous effects of migration on employment rates. Clockwise from the top left, the figures plot the change in employment for (a) women, (b) men, (c) movers who have a sibling who already moved, (d) movers who did not have a sibling who moved. All effects are normalised relative to β_{-2}, 95 per cent confidence intervals are shown.

carpenters, material handling, docking and freight handlers, and transport equipment operators. A fuller analysis of occupational upgrading will be presented in the following section, however these preliminary observations suggest that the differing returns between men and women may be due to differences in the occupations available to them in the urban area.

The labour market effects of urban migration likely depend on whether an individual has an existing social network or prior knowledge about the process of migrating. For this reason, individuals who are the first or only one of their siblings to move may experience lower employment effects than individuals who have a sibling who has already migrated. Alternatively, it could be that the first sibling leaves to smooth income ex-ante, while the subsequent sibling migrants are responding to income shocks ex-post (Gröger and Zylberberg, 2016). In the bottom panel of Figure 3.2 I explore the difference between first and subsequent movers, in the left and right panels respectively. The results are also in columns 3 and 4 of Table 3.A.1. The labour market effect is, if anything, smaller for siblings who move after another sibling has already moved. Interestingly, however, none of the pre-event dummies are significant for siblings who migrate after another sibling, indicating less evidence of an Ashenfelter dip. This would be consistent with a model where individuals don't migrate because they lack knowledge about how to find work in the city (Bryan et al., 2014). The first movers only move when they are pushed to emigrate by negative individual-specific economic shocks, however subsequent movers can take advantage of the knowledge gained by their siblings so don't need a shock to push them.

Finally, urban economists have argued that one of sources of higher earnings in bigger cities are thicker labour markets, making job finding easier. In column 5 of Table 3.A.1 I show that individuals who describe their destination as a city rather than a town do indeed experience larger employment gains than those who emigrate to towns, but the gains are no more permanent.

3.4.4 *Occupational upgrading*

I now turn to occupational choice as an alternative measure of labour market performance and evaluate whether urban migration leads to occupational upgrading relative to stayers. Studying occupations is complicated by two factors. First, occupation is a categorical variable; I transform it into a continuous variable using average occupational earnings in the 1995 Intercensus Population Survey, conducted by BPS and available from IPUMS.[4] Second, occupation is only observed for employed individuals, yet being employed is itself endogenous and, as the previous section demonstrates, significantly affected by urban migration, at least in the short term. Estimates of the effect of urban migration on occupational earnings conditional on being employed may, therefore, suffer from selection bias.

In column 1 of Table 3.A.2, I estimate the effect of moving to the city on the log of occupational earnings, irrespective of whether the mover is still in the urban area in k. The effect is significantly positive for $k \in [0, 6]$, albeit sometimes only at the ten per cent level, though it is zero in the longer term. In column 2 I estimate the effect, conditioning this time on an individual still being in an urban location, and find that the effect is clearly positive throughout the horizon under study. The finding of positive occupational earnings effects is in contrast to Hicks et al. (2017), who find no effect of migrating on earnings when using retrospectively reported annual earnings in the IFLS data (only available in waves 1-3). Pulido and Święcki (2019) argue the null earnings result can be explained by recall bias attenuating the true effect. In my case, recall bias is less of a problem, since individuals only need describe the main activities of the job, which is then classified into a given two-digit occupation using a dictionary mapping activities to occupations.

Clearly there is no permanent effect of migration on occupational earnings, which explains the difference between columns 1 and 2. There may also be negative selection on occupational earnings into return migration. In this case, the true effect on occupational earnings will be less than the one reported in column 2, though it is not possible to say whether it will be greater or less than the effect in column 1. Conditioning the analysis on being employed may also create an additional type of selection bias, since the set of employed individuals in the city may have

[4] This implies that I do not allow for within-occupation earnings differences between rural and urban locations. If occupational earnings are higher in the urban location, I will under-estimate the true earnings effect.

different unobservable characteristics, on average, to the set of employed individuals in the countryside.

To better understand occupational transitions, and explore any possible selection into employment, Table 3.5 plots the transition probabilities for all one-digit occupations for rural-urban migrants. The top panel reports transitions from $k = -1$ to $k = 0$, i.e. when moving, while the bottom panel reports transitions once in the city, conditional on not moving. Each base year occupation corresponds to a column. There is clearly a lot of persistence in occupations, the probability of remaining in the same occupation is always over 70 per cent when moving and over 80 per cent after moving. Furthermore, individuals who stay in the same one-digit occupation have a less than one per cent chance of moving to a two-digit occupation with higher average earnings. Focusing more specifically on selection into employment, the most common one-digit occupation at the move for individuals who were previously unemployed is low-skilled service work. Within this group, 53 per cent work in the two-digit occupation maids and related housekeeping service workers, the lowest-paid occupation. More generally, however, 29 per cent of unemployed movers who find a job do so in one of the ten worst-paid occupations, compared with 48 per cent of their employed stayer siblings in $k = -1$. This indicates that urban migration might lead to some hollowing out of the employment distribution relative to stayers. On average it offers the opportunity to find employment in better-paid occupations than stayers, however some movers do go from unemployment to working at the very bottom of the occupation distribution. The net effect of migration on occupation would then be ambiguous.

To avoid possible bias from selection into employment, I estimate the effect of urban migration on the probability of being in an occupation over a given threshold, where unemployment is the lowest-ranked occupation. There are 92 occupations in my data. The median employed stayer is in occupation 13 in $k = -1$, labourers not elsewhere classified, and the mean employed stayer is in occupation 21, housekeeping and related

Table 3.5: Occupational transition matrix for urban migrants

	Technical	Prof.	Clerical	Sales	Low-skill S.	Agric.	Unskilled P.	Skilled P.	TCP	Unemployed
$k = -1$ to $k = 0$										
Technical	80.00	5.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.76
Prof.	0.00	84.21	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.13
Clerical	0.00	0.00	84.62	1.25	1.45	0.50	1.20	4.00	0.99	2.77
Sales	0.00	0.00	0.00	71.25	2.90	4.02	4.82	4.00	1.98	7.19
Low-skill S.	0.00	0.00	0.00	6.25	82.61	2.01	3.61	0.00	3.96	13.75
Agric.	0.00	0.00	0.00	5.00	1.45	72.86	1.20	0.00	1.98	2.02
Unskilled P.	0.00	0.00	0.00	0.00	0.00	3.02	73.49	0.00	0.99	3.91
Skilled P.	0.00	0.00	0.00	0.00	1.45	3.02	1.20	84.00	0.99	4.16
TCP	0.00	0.00	3.85	0.00	0.00	4.02	3.61	0.00	85.15	6.68
Unemployed	20.00	10.53	11.54	16.25	10.14	10.55	10.84	8.00	2.97	57.63
Total	10	19	26	80	69	199	83	25	101	793
$k > 0$										
Technical	82.22	0.00	0.00	0.52	0.39	0.00	0.00	0.57	0.00	0.40
Prof.	0.00	79.49	0.00	0.79	0.00	0.00	0.46	0.57	0.25	0.97
Clerical	4.44	2.56	82.84	1.31	0.59	0.52	0.46	1.14	1.23	1.13
Sales	0.00	2.56	3.73	81.41	1.77	2.36	3.20	4.55	3.19	2.82
Low-skill S.	8.89	1.28	0.00	3.14	87.80	3.15	2.74	0.57	2.21	3.78
Agric.	0.00	0.00	1.49	1.31	1.97	83.20	2.74	1.70	1.47	1.21
Unskilled P.	0.00	2.56	0.75	1.05	1.57	1.05	80.82	1.14	1.23	1.53
Skilled P.	0.00	2.56	2.24	0.26	0.59	0.79	0.91	83.52	0.98	0.64
TCP	0.00	1.28	3.73	3.66	1.18	4.72	2.28	2.84	85.29	2.49
Unemployed	4.44	7.69	5.22	6.54	4.13	4.20	6.39	3.41	4.17	85.04
Total	45	78	134	382	508	381	219	176	408	1243

Note: The top panel shows transitions from event time $k = -1$ to event time $k = 0$, the bottom panel shows transitions for $k > 0$ conditional on not having moved again. Base year occupation is reported in columns. Services are abbreviated as S., Production work as P., and TCP is Transport, Construction, and Printing.

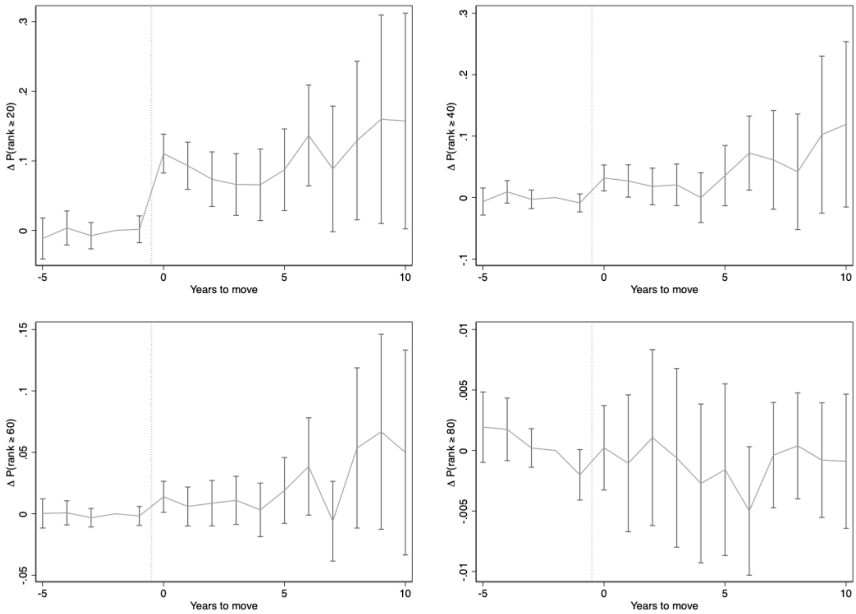
service supervisors while the median employed mover is in occupation 19 in $k = -1$, cabinet makers and related wood makers, and the mean employed mover is in occupation 23, wood preparation workers and paper makers. Given that urban migration only appears to have a longer-term effect on occupational earnings for individuals who stay in the urban location, I focus directly on estimates conditional on still being in the urban location. I will return to the question of selection into return migration with respect to occupational outcomes in Section 3.5.3.

In Figure 3.3 I report the effect of urban migration on the probability that the occupational rank is above a given threshold: 20, 40, 60, or 80, conditional on still being in the urban location. There is a significant effect of urban migration on the probability that the occupational rank is greater than 20, which increases by 12 percentage points in the year of the move, declines somewhat thereafter, then increases again, although the increase in $k \geq 5$ is very imprecisely estimated. There is also no evidence of any pre-trend or Ashenfelter dip in the occupational rank, unlike in the employment rate. Other thresholds see smaller, and not always significant, increases in the probability that the occupational rank is above them. There is also some evidence of an increasing trend for the thresholds 40 and 60, though again this is not precisely estimated.

The effect on the probability that the occupational rank is greater than 20 is larger than the effect on the employment rate, conditional on not re-migrating, previously reported in column 3 of Table 3.4, particularly for $k \geq 2$. This suggests that positive occupational earnings effect in column 2 of Table 3.A.2 is not reducible to selection into employment, particularly at longer horizons when the employment effect is zero. Estimates of the effect of moving on the occupational rank including return movers, reported in columns 3 and 4 of Table 3.A.2 for the thresholds 20 and 40, are smaller than the effect of non-return movers, consistent with previous evidence on occupational earnings.

A possible alternative explanation of the observed positive, and possibly increasing, long-run effect of moving on occupational earnings,

Figure 3.3: Occupational earnings effect of urban migration



Notes: Effect of urban migration on occupational rank, where unemployment is ranked zero, conditional on being in the urban location. All effects are normalised relative to β_{-2}, 95 per cent confidence intervals are shown.

is that some individuals move for educational reasons, and acquiring human capital will increase their earnings in the longer run, with returns potentially growing over time. In columns 4 and 5 of Table 3.A.2 I therefore add a further restriction, looking only at individuals who stated that they moved for reasons other than education and their siblings. As when imposing a similar restriction on employment, the effect is, if anything, larger, particularly at longer time horizons, consistent with those who move for educational purposes not working and potentially only starting to climb the job ladder later in event time.

Urban migrants do appear to experience occupational upgrading relative to their siblings. However, while this suggests that the return to migration is positive, it does not tell us how well migrants do compared to urban natives, and whether they experience any kind of labour market penalty.

3.5 LABOUR MARKET ASSIMILATION OF URBAN MIGRANTS

The previous section evaluated the return to urban migration, understood as the labour market performance of movers relative to a counterfactual situation in which the individual didn't move. In this section I turn to the question of whether urban migrants suffer any transient or permanent penalty in urban labour markets relative to a comparable set of urban residents.

3.5.1 *Matching procedure and empirical approach*

I use a matching procedure to identify a set of urban residents who are observationally similar, and hence comparable, to urban migrants. As before, I will define the year of an immigrant's move as the event year d, and define event time as $k = t - d$, where t refers to calendar time. The matching procedure uses a combination of time-invariant and pre-migration characteristics and proceeds in two steps.

The set of urban migrants are all individuals who report being born in a village, growing up in a village, and first appear in the data in a village but subsequently move to a town or city. I define the migration event to be the year in which they first migrated to the urban area. The set of match candidates are all individuals who were born in an urban area, resided there at age twelve, and first appear in the data in an urban area. I restrict my data to individuals aged 12 to 60, as above, and I impose a further data availability restriction. I only consider movers whom I observe in my dataset in at least $d - 1$, d, and $d + 8$. To be to be considered as a potential match for individuals who move in $d = t$, data must also be available for urban residents in $d-1$, d, and $d+8$.

In the first step of the matching procedure, I match all urban migrants in the event year d with all urban residents of the same gender present in the dataset in d. In the second step, I use Mahalanobis matching within year-gender cells to pair movers with a comparable urban resident. Mahalanobis matching minimises the Euclidean distance in normalised covariate space between paired observations, giving more-equal weight to each of the covariates than propensity score matching. I use employment in $d-1$, standardised Raven score, age at which the Raven test was taken, an interaction between standardised score and test age, and dummy variables for age and most recently observed 1-digit occupation in $d-1$ (which is zero if the individual has never worked before moving) as matching variables.[5] An urban resident may be chosen as a comparison observation multiple times in different event years d, but only once within a given event year.

Table 3.6 reports summary statistics for the urban migrants and comparison urban residents in $k = -1$. By construction, the share of women in the two groups is the same, at 53.6 per cent and average age is similarly balanced, at 21.65 years. Urban migrants are on average 0.12 lower in the birth order, score 0.07 of a standard deviation lower on the Raven test, are two percentage points more likely to be employed although, conditional on being employed, they are on average 1.5 ranks lower in the occupational distribution. Interestingly, the educational distribution is more dispersed for migrants. While urban residents are five percentage points more likely to complete senior high school, urban migrants are both more likely to have completed more than senior high and less than senior high. A chi-squared test of independence

[5]Note that including categorical variables via sets of dummy variables in the covariate set does cause these variables to be weighted more heavily; the standardised distance between two observations on a set of dummy variables defined on a categorical variable is bounded above by $\sqrt{2}\frac{n^2}{n-1}$, where n is the size of the dataset, which is independent of the number of categories.

between migrant status and highest completed education rejects the null of independence at the five per cent level (p-value = 0.04). All in all, however, these results suggest that the matching procedure identifies a comparison group of urban natives who are broadly comparable to urban migrants.

To estimate the difference in outcomes between migrants and urban natives, I will focus on the differences in average outcomes between urban migrants and natives in event year k. I will typically estimate these from the following specification:

$$y_{ik} = \alpha_k + \sum_{s=-5}^{10} \beta_s \times \mathbf{1}(mover_i = 1) \times \mathbf{1}(k = s) + \alpha_{t(ki)} + \alpha_{age_{ki}} + \varepsilon_{ik}. \quad (3.2)$$

I do not include an individual fixed effect, as I did in the sibling event study, only age and calendar year fixed effect, since level differences between urban migrants and natives are of direct interest. The differences in year k therefore no longer need to be normalised relative to a base year, β_k measures the absolute difference in employment rates in year k between migrants and natives, conditional on controls. It is also important to note that urban natives are treated as a reference group for the outcomes of urban migrants, but there is no sense in which they represent a possible counterfactual outcome for migrants had they not moved. The differences estimated by the coefficients β_s in Equation (3.2) are purely descriptive.

3.5.2 *Employment results*

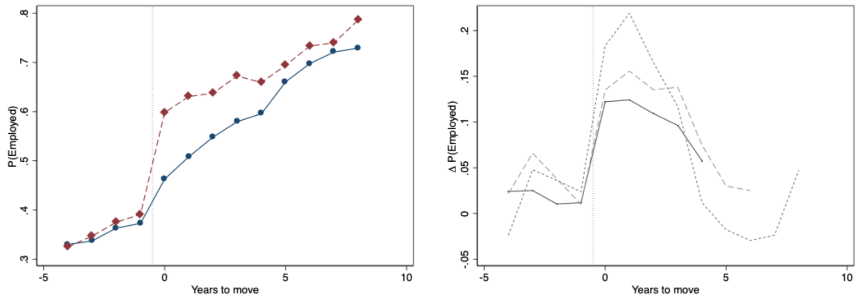
I first present descriptive evidence on the difference in employment rates between migrants and natives. In the left panel of Figure 3.4, I plot employment rates for urban migrants who are in the city in year k, excluding any episodes of return migration, and for their matched urban natives. That is, I include all migrant-native pairs where the immigrant is in the city in year k or where $k \leq 0$. This is similar to what Lubotsky

Table 3.6: Migrants and matched urban natives in $d-1$

	(1) Native	(2) Migrant	(3) Difference
P(Female)	0.54	0.54	0.00
	(0.50)	(0.50)	(0.02)
Age	21.65	21.65	-0.00
	(7.20)	(7.22)	(0.22)
Birth order	2.06	1.94	-0.11*
	(1.51)	(1.36)	(0.04)
Cognitive score	0.17	0.10	-0.07*
	(0.91)	(0.99)	(0.03)
P(Employed)	0.37	0.39	0.02
	(0.48)	(0.49)	(0.01)
Rank	27.07	25.50	-1.57
	(20.43)	(21.39)	(1.04)
No formal schooling	0.06	0.06	0.01
	(0.24)	(0.25)	(0.01)
Elementary school	0.20	0.22	0.01
	(0.40)	(0.41)	(0.01)
Junior high school	0.22	0.23	0.01
	(0.41)	(0.42)	(0.01)
Senior high school	0.35	0.31	-0.05**
	(0.48)	(0.46)	(0.01)
College (D1,D2,D3)	0.07	0.08	0.01
	(0.25)	(0.27)	(0.01)
University (BA, MA, PhD)	0.09	0.10	0.01
	(0.29)	(0.30)	(0.01)
Observations	2,109	2,109	4,218

Note: Information drawn from the employment and migration history modules of the IFLS, waves 1-5. Compares individuals born and raised in a village who move with a matched urban resident in year $k=-1$. + p<0.1, * p<0.05, ** p<0.01.

Figure 3.4: Employment of urban migrants relative to natives



Notes: Employment assimilation profile with respect to urban natives. In the left hand panel, urban migrants are plotted as red diamonds, urban natives as blue circles. In the right hand panel I plot average employment rates for urban natives minus the employment rates for urban migrants, conditional on the urban migrant spending at least 4, 6, or 8 years in the city. See text for details of matching procedure identifying comparison group of urban natives.

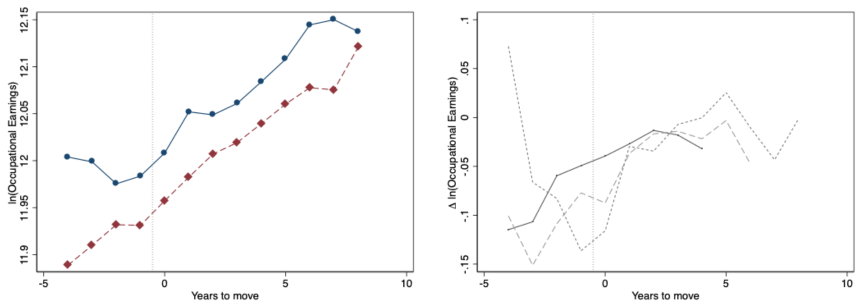
(2007) calls an "adjusted" date of arrival, i.e. one that ignores return migration spells, so long as an individual is still in the country or, in this case, city. This corresponds to 49 per cent of post-migration observations of migrant-native pairs. The year of migration is the year of original migration, regardless of whether there is an intervening return migration. The two groups appear to be on very similar trends before the migrant moves, after which the migrants have higher employment rates than the urban natives. The gap then contracts, though remains positive over the period considered. There is also a clear upward trend, reflecting the effect of age.

The descriptive results do not tell us what the pattern of labour market assimilation of urban migrants is, however, since return migration changes the set of included urban migrants across event years $k > 0$. As I documented in Section 3.3, return migration takes place at a rate of around 20 per cent per year, such that very few of the original migrants remain after more than five years. Return migration is, no doubt, endogenous and return migrants are likely to be negatively selected on realised labour market outcomes (for evidence in the case of international migration, see Lubotsky, 2007). In this case, conditioning the analysis on an individual still being in the city k years after migration will give an unbiased estimate of labour market assimilation for this subgroup of migrants in year k. However, since return migration is changing the composition of the set of urban migrants from year to year, the differences in employment rates are not comparable across years.

There is no clear-cut solution to the problem of selective return migration. I proceed by estimating differences in average employment rates between migrants and natives for migrant-native pairs where the migrant is still in the urban location after four, six, or eight years, constituting respectively 32, 19, and 8 per cent of the original sample. I plot the average employment rate of urban migrants minus the average employment rate of urban natives in the right panel of Figure 3.4 for the three groups. As when considering the time-varying sample, migrants are initially more likely to be employed than natives, and the difference increases as one focuses on groups of migrants who stay in the city longer. This clearly shows that, as in international migration, negative labour market outcomes lead to return migration for urban migrants. Furthermore, restricting attention to migrants who stay eight years, the difference with comparable natives becomes negligible in the longer term. This suggests that urban migrants tend to move with a job offer in hand, but once that job is over, they are no more likely than natives to be employed in the urban area.

Finally, in Table 3.A.3 I report the same difference in employment rates, this time controlling for age and year dummies, as well as the measured cognitive score of each individual. The included covariates do not materially affect the estimated differences in employment rates, reflecting the fact that the matching procedure led to a sample that is quite balanced with respect to these characteristics.

Figure 3.5: Occupational earnings of urban migrants relative to urban natives



Notes: Log occupational earnings assimilation profile with respect to urban natives. In the left hand panel, urban migrants are plotted as red diamonds, urban natives as blue circles. In the right hand panel I plot the average log occupational earnings for urban natives minus the average log occupational earnings for urban migrants, conditional on the urban migrant spending at least 4, 6, or 8 years in the city. See text for details of matching procedure identifying comparison group of urban natives.

3.5.3 *Occupation results*

I now turn to analysing occupational differences between urban migrants and natives by years since migration, conditional on employment. In the left panel of Figure 3.5, I plot log occupational earnings for both urban migrants and urban natives, again conditional on the urban migrant being in the urban location in period k. While there is evidence of some convergence *before* migration, on the whole the average earnings gap between migrants and natives is quite stable at around five log points over the period considered. Since the two groups are very similar in education, cognitive score, gender, and age at migration, this level difference might be evidence that human capital acquired in the rural area is less valuable than the same quantity of human capital acquired in the urban area.

As in the analysis of employment differences between migrants and natives, selective return migration and back-and-forth migration are a potential source of bias, or at least imply that the results can only be interpreted as representative for immigrants who stay in the country at least up until period k. To better understand selective outmigration, I again plot the difference between the average log occupational earnings of urban migrants and natives, conditional on employment and on not re-migrating after four, six, or eight years, in the right panel of Figure 3.5. Here there is stronger evidence of occupational convergence on the part of migrants, as the differences in log average earnings are relatively small, though somewhat volatile across years, for $k \geq 3$. Regression estimates controlling for age, year, and cognitive score, presented in columns 1-3 of Table 3.A.4, confirm that, while urban migrants are in lower-earning occupations before moving, there is no significant difference in occupational earnings between urban migrants and natives once the former are in the urban location.

As in the sibling event studies reported in Section 3.4.4, however, selection into employment might bias comparisons of employed urban migrants and urban natives. If less able urban migrants have more trouble finding employment in the city than similarly skilled urban residents, perhaps because they have less dense job-finding networks than the natives, estimated occupational differences between urban migrants and natives are likely to be biased upward. I address this possibility in columns 4-6 of Table 3.A.4, where I report differences between urban migrants and urban natives in the probability of being in an occupation above the twentieth rank, conditional on age, year, and cognitive score. I focus on the same three subgroups; i.e. individuals who have not re-migrated after four, six, and eight years. Unemployed individuals are assigned the lowest occupational rank. The probability of holding a job at or above the twentieth rank is around seven percentage points higher for migrants who stay 4-6 years than for comparable natives in $k = 0$, and 10.6 percentage points higher for migrants who stay 8 years. Comparing these differences to the difference between migrants' and urban natives' employment rates, one may note that the initial occupational difference is 5.5-10 percentage points lower for all groups than the corresponding employment difference, in columns 2-4 of Table 3.A.3. This indicates that an important part of the employment difference between migrants and urban natives comes from migrants taking jobs in the lower part of the occupational distribution.

Turning to the differences in subsequent periods, one can observe a similar evolution of the difference in the probability of holding a job in an occupation above the twentieth rank, which increases in $k = 1$ and decreases thereafter. There is no evidence of the difference in the probability of working in a higher-ranked occupation overtaking the difference in employment probabilities. In the longer run, the difference between migrants and natives in occupational rank is statistically indistinguishable from zero. There does not appear to be any evidence of migrants experiencing an occupational penalty relative to natives in the longer run, conditional on staying in the urban area.

3.6 CONCLUSION

In summary, moving to the city leads to transient employment gains for rural-urban migrants and more long-term occupational upgrading relative to sibling stayers. Sample selection issues complicate the interpretation of these results, since observed migrants are likely to be positively selected, and migrants who do worse than they expected are more likely to remigrate. Such positive selection seems unlikely to entirely explain the observed results; however, a thorough answer to the question is a subject for future research. Comparing urban migrants to similar urban natives also indicates that urban migrants do not under-perform relative to comparable natives.

These findings help shed some more light on the reasons why more migrants do not move to the city. The existence of large, if transient, earnings effects, coupled with evidence of occupational upgrading suggests that the barriers to migration are larger than simple comparisons of wages conditional on employment would suggest. This is consistent with recent evidence that moving costs (Bryan and Morten, 2019; Pulido and Święcki, 2019) and a preference for living in one's birthplace (Zerecero, 2020) are important barriers to an efficient allocation of individuals across space.

However, a precise accounting for the relative importance of migration costs and home bias, and how efficiently sorted individuals are, would require re-evaluating the estimates presented here through a formal model that accounts for the possibility of being unemployed, a productive avenue for future research. Another productive avenue for research will be to extend the research design used here to other developing countries, to understand how general the patterns identified are.

3

APPENDIX

Appendix 3.A supplementary tables

Table 3.A.1: Heterogeneity in employment outcomes

	(1) Female	(2) Male	(3) First mover	(4) Later mover	(5) Mig. to city
Movers × k = -5	0.020	0.007	0.004	0.028	0.043
	(0.028)	(0.026)	(0.022)	(0.039)	(0.030)
Movers × k = -4	0.041^{+}	0.014	0.020	0.044	0.036
	(0.023)	(0.023)	(0.018)	(0.034)	(0.026)
Movers × k = -3	0.013	-0.010	-0.000	0.007	0.014
	(0.019)	(0.019)	(0.015)	(0.026)	(0.021)
Movers × k = -1	-0.037*	-0.005	-0.026^{+}	-0.007	0.010
	(0.018)	(0.017)	(0.014)	(0.025)	(0.020)
Movers × k = 0	0.127**	0.119**	0.124**	0.116**	0.194**
	(0.026)	(0.022)	(0.020)	(0.033)	(0.027)
Movers × k = 1	0.075*	0.097**	0.076**	0.113**	0.172**
	(0.029)	(0.024)	(0.022)	(0.040)	(0.031)
Movers × k = 2	0.034	0.062*	0.045^{+}	0.058	0.122**
	(0.031)	(0.025)	(0.023)	(0.041)	(0.032)
Movers × k = 3	-0.004	0.051*	0.022	0.022	0.064^{+}
	(0.033)	(0.026)	(0.024)	(0.042)	(0.034)
Movers × k = 4	-0.026	0.017	0.009	-0.065	0.030
	(0.036)	(0.029)	(0.025)	(0.044)	(0.035)
Movers × k = 5	-0.028	0.045	0.016	-0.034	0.044
	(0.035)	(0.028)	(0.025)	(0.046)	(0.034)
Movers × k = 6	0.000	0.043	0.028	-0.026	0.061^{+}
	(0.037)	(0.029)	(0.026)	(0.049)	(0.035)
Movers × k = 7	-0.022	0.066*	0.027	-0.035	0.062
	(0.039)	(0.030)	(0.027)	(0.051)	(0.038)
Movers × k = 8	-0.036	0.075*	0.007	0.012	0.014
	(0.042)	(0.030)	(0.028)	(0.055)	(0.039)
Movers × k = 9	-0.044	0.063*	0.004	-0.048	0.004
	(0.043)	(0.029)	(0.028)	(0.057)	(0.040)
Movers × k = 10	-0.011	0.083**	0.017	0.050	0.046
	(0.046)	(0.031)	(0.029)	(0.067)	(0.043)
N	20998	20222	31821	9398	18161
Clusters	910	923	1091	257	542
R^2	0.48	0.60	0.55	0.57	0.55

Note: Column one estimates the effect for female movers, column 2 for male movers, column 3 for the first sibling I observe moving, column 4 for subsequent siblings I observe moving, column 5 for individuals who move to a city rather than a town. Coefficients are normalised relative to β_{-2}, which is omitted, standard errors are clustered by sibling group. + p<0.1, * p<0.05, ** p<0.01.

Table 3.A.2: Occupational of movers relative to sibling stayers

	All obs.	Urban only	All obs.		Urban + no educ.	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Earn.)	ln(Earn.)	P(rk $\geq$ 20)	P(rk $\geq$ 40)	P(rk $\geq$ 20)	P(rk $\geq$ 40)
Movers $\times$ k = -5	-0.002	-0.004	-0.014	-0.005	-0.010	-0.008
	(0.020)	(0.020)	(0.015)	(0.011)	(0.016)	(0.012)
Movers $\times$ k = -4	0.006	0.015	0.001	0.009	0.007	0.010
	(0.016)	(0.015)	(0.013)	(0.009)	(0.013)	(0.010)
Movers $\times$ k = -3	-0.018	-0.007	-0.008	-0.003	-0.008	-0.002
	(0.012)	(0.012)	(0.010)	(0.008)	(0.010)	(0.008)
Movers $\times$ k = -1	0.013	0.006	0.005	-0.008	0.009	-0.005
	(0.012)	(0.011)	(0.010)	(0.007)	(0.010)	(0.008)
Movers $\times$ k = 0	0.063**	0.047**	0.113**	0.033**	0.117**	0.036**
	(0.016)	(0.015)	(0.014)	(0.011)	(0.015)	(0.011)
Movers $\times$ k = 1	0.048*	0.047*	0.089**	0.022+	0.101**	0.036*
	(0.019)	(0.018)	(0.016)	(0.013)	(0.018)	(0.014)
Movers $\times$ k = 2	0.040+	0.043+	0.077**	0.017	0.083**	0.026+
	(0.021)	(0.022)	(0.018)	(0.014)	(0.021)	(0.016)
Movers $\times$ k = 3	0.043+	0.053*	0.066**	0.019	0.076**	0.028
	(0.022)	(0.025)	(0.019)	(0.015)	(0.023)	(0.018)
Movers $\times$ k = 4	0.042+	0.060*	0.059**	0.006	0.076**	0.009
	(0.024)	(0.028)	(0.021)	(0.017)	(0.027)	(0.021)
Movers $\times$ k = 5	0.046+	0.083**	0.068**	0.022	0.100**	0.040
	(0.025)	(0.031)	(0.022)	(0.018)	(0.031)	(0.026)
Movers $\times$ k = 6	0.047+	0.127**	0.088**	0.034+	0.153**	0.078*
	(0.027)	(0.038)	(0.023)	(0.019)	(0.038)	(0.031)
Movers $\times$ k = 7	0.022	0.092*	0.076**	0.030	0.100*	0.068+
	(0.027)	(0.045)	(0.025)	(0.020)	(0.047)	(0.041)
Movers $\times$ k = 8	0.017	0.108*	0.069**	0.023	0.149*	0.052
	(0.028)	(0.047)	(0.026)	(0.021)	(0.059)	(0.048)
Movers $\times$ k = 9	0.011	0.139*	0.045+	0.030	0.173*	0.113+
	(0.029)	(0.065)	(0.027)	(0.022)	(0.077)	(0.066)
Movers $\times$ k = 10	0.007	0.101	0.062*	0.048*	0.177*	0.132+
	(0.031)	(0.069)	(0.028)	(0.024)	(0.080)	(0.069)
N	24693	20736	41078	41078	31250	31250
Clusters	1057	1038	1091	1091	1091	1091
R^2	0.73	0.75	0.56	0.50	0.59	0.53

Note: Columns 1, 3, and 4 report effects for all individuals, i.e. without conditioning on a mover still being in the urban location. Columns 2, 5, and 6 report effects conditional on the movers still being in the urban location. Columns 5 and 6 also condition on the reason for the move not being education. Coefficients are normalised relative to β_{-2}, which is omitted, standard errors are clustered by sibling group. + p<0.1, * p<0.05, ** p<0.01.

Table 3.A.3: Employment rates of migrants relative to matched urban natives

	(1) All urban	(2) Duration $\geq$ 4	(3) Duration $\geq$ 6	(4) Duration $\geq$ 8
Migrant $\times$ k = -4	0.001	0.027	0.013	-0.037
	(0.014)	(0.027)	(0.036)	(0.049)
Migrant $\times$ k = -3	0.013	0.026	0.061^{+}	0.032
	(0.014)	(0.025)	(0.035)	(0.049)
Migrant $\times$ k = -1	0.021	0.014	0.014	0.027
	(0.013)	(0.023)	(0.031)	(0.047)
Migrant $\times$ $k = 0$	0.138^{**}	0.124^{**}	0.139^{**}	0.187^{**}
	(0.014)	(0.025)	(0.033)	(0.053)
Migrant $\times$ $k = 1$	0.125^{**}	0.127^{**}	0.160^{**}	0.223^{**}
	(0.016)	(0.026)	(0.034)	(0.052)
Migrant $\times$ $k = 2$	0.093^{**}	0.112^{**}	0.140^{**}	0.169^{**}
	(0.018)	(0.025)	(0.033)	(0.052)
Migrant $\times$ $k = 3$	0.095^{**}	0.099^{**}	0.141^{**}	0.117^{*}
	(0.021)	(0.025)	(0.032)	(0.050)
Migrant $\times$ $k = 4$	0.066^{**}	0.060^{*}	0.079^{*}	0.014
	(0.023)	(0.026)	(0.032)	(0.051)
Migrant $\times$ $k = 5$	0.036		0.034	-0.015
	(0.024)		(0.032)	(0.050)
Migrant $\times$ $k = 6$	0.039		0.029	-0.028
	(0.025)		(0.031)	(0.048)
Migrant $\times$ $k = 7$	0.022			-0.022
	(0.027)			(0.048)
Migrant $\times$ $k = 8$	0.060^{*}			0.048
	(0.029)			(0.048)
N	32742	11288	8224	4141
Clusters	4218	1354	800	338
R^2	0.24	0.20	0.19	0.21

Note: Column one reports employment differences for Coefficients are normalised relative to β_{-2}, which is omitted. Duration refers to the minimum number of years an individual is observed in the urban location, the sample is truncated at the chosen threshold. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01.

Table 3.A.4: Occupation of migrants relative to urban natives

	ln(Earn.)			P(rk ≥ 20)		
	(1) Duration ≥ 4	(2) Duration ≥ 6	(3) Duration ≥ 8	(4) Duration ≥ 4	(5) Duration ≥ 6	(6) Duration ≥ 8
Migrant × k = -4	-0.103*	-0.101+	0.026	-0.014	-0.025	0.009
	(0.043)	(0.059)	(0.094)	(0.023)	(0.030)	(0.041)
Migrant × k = -3	-0.090*	-0.136*	-0.068	-0.012	-0.011	0.003
	(0.040)	(0.054)	(0.095)	(0.022)	(0.030)	(0.040)
Migrant × k = -1	-0.032	-0.062	-0.123	-0.003	-0.009	-0.013
	(0.036)	(0.047)	(0.074)	(0.021)	(0.028)	(0.039)
Migrant × $k = 0$	0.004	-0.038	-0.061	0.070**	0.058+	0.093+
	(0.030)	(0.041)	(0.063)	(0.024)	(0.032)	(0.050)
Migrant × $k = 1$	0.000	-0.006	0.003	0.085**	0.107**	0.177**
	(0.029)	(0.038)	(0.061)	(0.025)	(0.033)	(0.049)
Migrant × $k = 2$	0.002	-0.001	-0.029	0.083**	0.101**	0.134**
	(0.028)	(0.037)	(0.060)	(0.025)	(0.033)	(0.051)
Migrant × $k = 3$	-0.006	-0.002	-0.006	0.070**	0.101**	0.115*
	(0.028)	(0.037)	(0.058)	(0.026)	(0.034)	(0.052)
Migrant × $k = 4$	-0.026	-0.018	-0.010	0.041	0.062+	0.045
	(0.028)	(0.035)	(0.055)	(0.026)	(0.034)	(0.053)
Migrant × $k = 5$		0.003	0.023		0.039	0.034
		(0.035)	(0.054)		(0.035)	(0.053)
Migrant × $k = 6$		-0.036	-0.008		0.008	-0.003
		(0.034)	(0.052)		(0.035)	(0.053)
Migrant × $k = 7$			-0.042			-0.029
			(0.049)			(0.053)
Migrant × $k = 8$			0.001			0.042
			(0.048)			(0.053)
N	5602	4548	2306	11269	8213	4137
Clusters	1072	696	304	1354	800	338
R^2	0.09	0.08	0.10	0.14	0.14	0.18

Note: See text for details of matching procedure. Coefficients are normalised relative to β_{-2}, which is omitted. Duration refers to the minimum number of years an individual is observed in the urban location, the sample is truncated at the chosen threshold. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01.

BIBLIOGRAPHY

Abadie, A., Imbens, G. W., Athey, S., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *NBER Working Papers*, 24003.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.

Abramitzky, R., Boustan, L. P., and Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses : Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102(5):1832–1856.

Abramitzky, R., Boustan, L. P., and Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, 122(3):467–506.

Aigner, D. J. and Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, 30(2):175–187.

Albelda, R. P. (1986). Occupational segregation by race and gender, 1958-1981. *ILR Review*, 39(3):404–411.

Aldén, L., Hammerstedt, M., and Neuman, E. (2015). Ethnic Segregation, Tipping Behavior, and Native Residential Mobility. *International Migration Review*, 49(1):36–69.

Almond, D. and Currie, J. (2011). Human Capital Development before Age Five. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4B, pages 1315–1486. Elsevier B.V.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on Observed and Unobserved Variables : Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184.

Andersson, F., García-Pérez, M., Haltiwanger, J., McCue, K., and Sanders, S. (2014). Workplace Concentration of Immigrants. *Demography*, 51(6):2281–2306.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62(6):1383–1414.

Andrews, I., Kitagawa, T., and McCloskey, A. (2020). Inference After Estimation of Breaks. *Journal of Econometrics*, Forthcomin.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricists Companion.* Princeton University Press, Princeton, N.J. :.

Ansala, L., Åslund, O., and Sarvimäki, M. (2021). Immigration history, entry jobs, and the labor market integration of immigrants. *Journal of Economic Geography.*

Arellano-Bover, J. (2020). Career Consequences of Firm Heterogeneity for Young Workers : First Job and Firm Size. *IZA Discussion Papers*, 12969.

Åslund, O., Hensvik, L., and Skans, O. N. (2014). Seeking Similarity: How Immigrants and Natives Manage in the Labor Market. *Journal of Labor Economics*, 32(3):405–441.

Åslund, O. and Skans, O. N. (2010). Will I see you at work: Ethnic Workplace Segregation in Sweden, 1985-2002,. *ILR Review*, 63(3):471–493.

Aydemir, A. and Skuterud, M. (2008). The immigrant wage differential within and across establishments. *ILR Review*, 61(3):334–352.

Azlor, L., Damm, A. P., and Schultz-Nielsen, M. L. (2020). Local labour demand and immigrant employment. *Labour Economics*, 63:101808.

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

Barth, E., Bratsberg, B., and Raaum, O. (2012). Immigrant wage profiles within and between establishments. *Labour Economics*, 19(4):541–556.

Battisti, M., Peri, G., and Romiti, A. (2018). Dynamic effects of co-ethnic networks on immigrants' economic success. *CESifo Working Paper*, 7084.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *Review of Economic Studies*, 79(1):128–161.

Becker, G. S. (1957). *The economics of discrimination: an economic view of racial discrimination*. University of Chicago.

Becker, G. S. and Murphy, K. M. (2000). *Social Economics: Market behaviour in a social environment*. Harvard University Press.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. B. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., and Hansen, C. B. (2013). Inference for High-Dimensional Sparse Econometric Models. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Advances in Economics and Econometrics: Tenth World Congress, Vol. 3*, pages 245–95. Cambridge University Press, Cambridge, UK.

Belloni, A., Chernozhukov, V., and Hansen, C. B. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *Review of Economic Studies*, 81(2):608–650.

Bevis, L., Kim, K., and Guerena, D. (2020). Soils and South Asian Stunting : Soil zinc deficiency drives child stunting in Nepal.

Böhlmark, A. and Willén, A. (2020). Tipping and the Effects of Segregation. *American Economic Journal: Applied Economics*, 12(1):318–347.

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3(4):463–489.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, 85(3):365–390.

Borjas, G. J. (1998). To Ghetto or Not to Ghetto : Ethnicity and Residential Segregation. *Journal of Urban Economics*, 44:228–253.

Borusyak, K. and Jaravel, X. (2017). Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume.

Boucher, V. and Goussé, M. (2019). Wage dynamics and peer referrals. *Review of Economic Dynamics*, 31:1–23.

Brücker, H., Kroh, M., Bartsch, S., Goebel, J., Kühne, S., Liebau, E., Trübswetter, P., Tucci, I., and Schupp, J. (2013). The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. *SOEP Survey Papers*, 216.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a Profitable Technology : The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82(5):1671–1748.

Bryan, G. and Morten, M. (2019). The Aggregate Productivity Effects of Internal Migration : Evidence from Indonesia. *Journal of Political Economy*, 127(5):2229–2268.

Burdett, K. and Mortensen, D. T. (1998). Wage Differentials, Employer Size, and Unemployment. *International Economic Review*, 39(2):257–273.

Caetano, G. and Maheshri, V. (2017). School segregation and the identification of tipping behavior. *Journal of Public Economics*, 148:115–135.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Calvó-Armengol, A. and Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, 94(3):426–454.

Card, D. (1999). The Causal Effect of Education on Earnings. In Ashenfelter, O. C. and Card, D., editors, *Handbook of Labour Economics*, volume 3A, pages 1801–1863. Elsevier B.V.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, 36(S1):S13–S70.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *Quarterly Journal of Economics*, 123(1):177–218.

Card, D., Mas, A., and Rothstein, J. (2011). Are Mixed Neighborhoods Always Unstable? Two-Sided and One-Sided Tipping. In Newburger, H. B., Birch, E. L., and Wachter, S. M., editors, *Neighborhood and life changes: How place matters in modern America*. University of Pennsylvania Press, Philadelphia.

Cattaneo, M. D., Jansson, M., and Ma, X. (2019). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 0(0):1–7.

Cattaneo, M. D. and Vazquez-Bare, G. (2016). The Choice of Neighborhood in Regression Discontinuity Designs. *Observational Studies*, 2:134–146.

Chernozhukov, V., Hansen, C. B., and Spindler, M. (2015). Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments. *American Economic Review, Papers and Proceedings*, 105(5):486–490.

Chiswick, B. R. (1978). The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy*, 86(5):897–921.

Cingano, F. and Rosolia, A. (2012). People I Know: Job Search and Social Networks. *Journal of Labor Economics*, 30(2):291–332.

Cutler, D. M. and Glaeser, E. L. (1997). Are Ghettos Good or Bad? *The Quarterly Journal of Economics*, 112(3):827–872.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The Rise and Decline of the American Ghetto. *Journal of Political Economy*, 107(3):455–506.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (2008). Is the melting pot still hot? Explaining the resurgence of immigrant segregation. *Review of Economics and Statistics*, 90(3):478–497.

Damm, A. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, 27(2):281–314.

De La Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *Review of Economic Studies*, 84(1):106–142.

Dustmann, C., Fasani, F., Meng, X., and Minale, L. (2017). Risk Attitudes and Household Migration Decisions. *IZA Discussion Papers*, 10603.

Dustmann, C., Glitz, A., Schönberg, U., and Brücker, H. (2016). Referral-based job search networks. *Review of Economic Studies*, 83(2):514–546.

Easterly, W. (2009). Empirics of strategic interdependence: The case of the racial tipping point. *B.E. Journal of Macroeconomics*, 9(1).

Eckstein, Z. and Weiss, Y. (2010). On the Wage Growth of Immigrants: Israel 1990-2000. *Journal of the European Economic Association*, 2(4):665–695.

Edin, P.-A., Fredriksson, P., and Åslund, O. (2003). Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *Quarterly Journal of Economics*, 118(1):329–357.

Eliason, M., Hensvik, L., Kramarz, F., and Nordstrom Skans, O. (2019). Social Connections and the Sorting of Workers to Firms. *IZA Discussion Papers*, 12323.

European Commission (2018). Integration of immigrants in the European Union. Technical Report Eurobarometer 469.

Farré, L. and Fasani, F. (2012). Media exposure and internal migration — Evidence from Indonesia. *Journal of Development Economics*, 102(C):48–61.

Galenianos, M. (2013). Review of Economic Dynamics Learning about match quality and the use of referrals. *Review of Economic Dynamics*, 16(4):668–690.

Gee, L. K., Jones, J., and Burke, M. (2017a). Social Networks and Labor Markets : How Strong Ties Relate to Job Finding on Facebook's Social Network. *Journal of Labor Economics*, 35(2):485–518.

Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., and Fowler, J. H. (2017b). The paradox of weak ties in 55 countries. *Journal of Economic Behavior and Organization*, 133:362–372.

Glitz, A. (2014). Ethnic segregation in Germany. *Labour Economics*, 29:28–40.

Glitz, A. (2017). Coworker networks in the labour market. *Labour Economics*, 44:218–230.

Glitz, A. and Vejlin, R. (2020). Learning through Coworker Referrals. *Review of Economic Dynamics*, Forthcomin.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: evidence from french grocery stores. *Quarterly Journal of Economics*, 132(3):1219–1260.

Goldin, C. (2014). A Pollution Theory of Discrimination : Male and Female Differences in Occupations and Earnings. In Boustan, L. P., Frydman, C., and Margo, R. A., editors, *Human Capital in History : The American Record*, pages 313–348. University of Chicago Press.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics*, 129(2):939–993.

Granovetter, M. (1995). *Getting a job: a study of contacts and carrers.* University of Chicago Press, 2 edition.

Gröger, A. (2021). Easy come , easy go ? Economic shocks , labor migration and the family left behind. *Journal of International Economics*, 128:1–22.

Gröger, A. and Zylberberg, Y. (2016). Internal Labor Migration as a Shock Coping Strategy: Evidence from a Typhoon. *American Economic Journal: Applied Economics*, 8(2):123–153.

Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, 64(2):413–430.

Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, 68(3):575–603.

Hansen, B. E. (2020). *Econometrics.* Unpublished Manuscript.

Harris, J. R. and Todaro, M. P. (1970). Migration, Unemployment and Developmnent : A Two-Sector Analysis. *American Economic Review*, 60(1):126–142.

Hellerstein, J. K. and Neumark, D. (2008). Workplace Segregation in the United States: Race, ethnicity, and skill. *The Review of Economics and Statistics*, 90(3):459–477.

Henderson, J. V. and Turner, M. A. (2020). Urbanization in the Developing World: Too Early or Too Slow? *Journal of Economic Perspectives*, 34(3):150–173.

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If?* Chapman and Hall, Boca Raton.

Hicks, J. H., Kleemans, M., Li, N. Y., and Miguel, E. (2017). Reevaluating Agricultural Productivity Gaps With Longitudinal Microdata. *NBER Working Paper*, 23253.

Higgs, B. Y. R. (1977). Firm-Specific Evidence on Racial Wage Differentials and Workforce Segregation. *American Economic Review*, 67(2):236–245.

Hjort, J. (2014). Ethnic Divisions and Production in Firms. *Quarterly Journal of Economics*, 129(4):1899–1946.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica*, 87(5):1439–1474.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality : Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4):1129–1179.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs : A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The Economic Consequences of Social-Network Structure. *Journal of Economic Literature*, 55(1):49–95.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the national hockey league. *Review of Economics and Statistics*, 95(March):302–314.

Kropp, P. and Schwengler, B. (2011). Abgrenzung von Arbeitsmarktregionen – ein Methodenvorschlag. *Raumforschung und Raumordnung*, 69(1):45–62.

Lagakos, D. (2020). Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities? *Journal of Economic Perspectives*, 34(3):174–192.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., and Waug (2020). Migration Costs and Observational Returns to Migration in the Developing World. *Journal of Monetary Economics*, 113:138–154.

Lagakos, D. and Waugh, M. E. (2013). Selection, Agriculture, and Cross-Country Productivity Differences. *American Economic Review*, 103(2):948–980.

Lazear, E. P. (1999a). Culture and Language. *Journal of Political Economy*, 107(S6):95–126.

Lazear, E. P. (1999b). Globalisation and the Market for Team-Mates Author(s): Edward P. Lazear Source: The Economic Journal, Vol. 109, No. 454, Conference Papers (Mar., 1999), pp. C15-C40 Published by:. *The Economic Journal*, 109(454):15–40.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(June):281–355.

Lubotsky, D. H. (2007). Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings. *Journal of Political Economy*, 115(5):820–867.

Manning, A. (2011). *Imperfect competition in the labor market*, volume 4. Elsevier B.V.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design : A density test. *Journal of Econometrics*, 142:698–714.

Mcpherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes : Toward an Economic Analysis. *American Economic Review*, 81(5):1408–1418.

Montgomery, J. D. (1992). Job Search and Network Composition : Implications of the Strength-Of-Weak-Ties Hypothesis. *American Sociological Review*, 57(5):586–596.

Morgan, S. L. and Winship, C. (2014). *Counterfactuals and Causal Inference*. Cambridge University Press, Cambridge, UK, 2 edition.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor marke. *Quarterly Journal of Economics*, 118(2):549–599.

Munshi, K. and Rosenzweig, M. (2016). Networks and Misallocation : Insurance , Migration , and the Rural-Urban Wage Gap. *American Economic Review*, 106(1):46–98.

Noiriel, G. (1988). *Le creuset français: histoire de l'immigration XIXe-XXe siècles*. Éditions du Seuil, Paris.

OECD (2020). International Migration Statistics.

Oster, E. (2019). Unobservable Selection and Coefficient Stability : Theory and Evidence. *Journal of Business and Economic Statistics*, 37(2):187–204.

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

Pan, J. (2015). Gender Segregation in Occupations : The Role of Tipping and Social Interactions. *Journal of Labor Economics*, 33(2):365–408.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2 edition.

Peri, G. and Sparber, C. (2009). Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics*, 1(3):135–169.

Porter, J. and Yu, P. (2015). Regression discontinuity designs with unknown discontinuity points : Testing and estimation. *Journal of Econometrics*, 189(1):132–147.

Pulido, J. and Święcki, T. (2019). Barriers to Mobility or Sorting ? Sources and Aggregate Implications of Income Gaps across Sectors in Indonesia. *Society for Economic Dynamics Meeting Paper*, 1298:1–70.

Rosen, S. (1986). The theory of equalising differences. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume I, pages 641–692. Elsevier, Amsterdam.

Rosenzweig, M. R. . and Stark, O. (1989). Consumption Smoothing , Migration , and Marriage : Evidence from Rural India. *Journal of Political Economy*, 97(4):905–926.

Sarvimäki, M. (2011). Assimilation to a Welfare State : Labor Market Performance and Use of Social Benefits by Immigrants to Finland. *Scandinavian Journal of Economics*, 113(3):665–688.

Sarvimäki, M., Uusitalo, R., and Jäntti, M. (2020). Habit Formation and the Misallocation of Labor : Evidence from Forced Migrations.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186.

Schelling, T. C. (1978). *Micromotives and Macrobehaviour*. Norton, New York.

Semyonov, M. and Glikman, A. (2009). Ethnic residential segregation, social contacts, and anti-minority attitudes in European societies. *European Sociological Review*, 25(6):693–708.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2019). Firming up inequality. *Quarterly Journal of Economics*, 134(1):1–50.

Sorkin, I. (2018). Ranking Firms Using Revealed Preference. *Quarterly Journal of Economics*, 133(3):1331–1393.

Stark, O. and Bloom, D. E. . (1985). The New Economics of Labor Migration. *American Economic Review, Papers and Proceedings*, 75(2):173–178.

Strauss, J., Witoelar, F., and Sikoki, B. (2016). The Fifth Wave of the Indonesia Family Life Survey : Overview and Field Report.

Young, A. (2013). Inequality, the Urban-Rural Gap, and Migration. *Quarterly Journal of Economics*, 128(4):1727–1785.

Zerecero, M. (2020). The Birthplace Premium.