Workplace segregation and the labour market performance of immigrants

Sébastien Willis

December 2020

Abstract

Segregation of immigrants across workplaces has been widely documented, however the consequences of segregation remain subject to conjecture. 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 0.15-0.2 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.

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 help explain 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 no evidence I know of linking individual workers' outcomes to the composition of their current or past workplaces.

In this paper, 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. The central identification claim of this paper will be that these 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.

Different theories make contrasting and sometimes contradictory predictions about the contemporaneous and the dynamic effects of workforce composition. For this reason, in the first part of the paper I systematically review the relevant theories and the empirical evidence supporting them, distinguishing in particular between short-term and long-term predictions. I also distinguish between predictions for the effect of the own-group share and the effect of the total immigrant share. Since different theories can make contradictory predictions about the effect of the composition of the first job an immigrant will hold, empirical evidence is necessary to help us understand which of the proposed theoretical mechanisms is most important.

In the second part of the paper, I study the association between the share of conationals when an immigrant starts out in her first job and the wages and employment rate of the immigrant over time. I find that starting out in an firm with a higher conational share is negatively associated with an immigrant's employment rate in the longer term; a ten percentage point increase in the initial conational share is associated on average with 1.5-2 percentage point lower employment rates six or more years after the first job. The long-term effects on earnings are also negative; a ten percentage point increase in the initial conational share is associated with a 3.5 log-point reduction in earnings after six years, conditional on being employed. The earnings effect, however, appears unevenly distributed across the distribution; the probability of earning at least 10,000 Euros decreases 2 percentage points with a 10 percentage point increase in the initial conational share, however there is no effect on the probability of earning at least 30,000 Euros. Importantly, these effects do not appear to be present for the share of immigrants who do not share the migrant's nationality, only for the conational share.

The bulk of the paper is then devoted to analysing the different forms of bias that might affect the estimated effects. First, I discuss the possibility that my results are biased by selection into return migration on the initial conational share. Selection into return migration on the variable of interest, or treatment, 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. This paper is the first to my knowledge to formally study the sign of the bias created by selection on the variable of interest into return migration. The first contribution of this paper is to show, under fairly general conditions, that the sign of selection bias depends on (i) the sign of the effect of the treatment of interest, in my case the initial conational share, on return migration; (ii) the sign of the effect of the outcome of interest, in my case the employment rate, on return migration; and (iii) the sign of direct effect of the treatment on the outcome, which my results suggest is negative.[1] I argue that *a priori* one might expect both effects (i) and (ii) to be negative, and show that the selection bias is therefore likely to be positive. In this case my estimated effects would underestimate the true (negative) effect. There is no evidence in my sample of a non-null covariance between the initial conational share and return migration, although this is likely due to my sample being too small to detect any relationship between the two.

Second, to assess to what extent the associational evidence might be explained by selection into high-conational share firms on remaining unobservable characteristics, I apply the test proposed by Oster (2019). The test formally assesses how strong selection on unobservables would have to be for the true causal effect to be zero, given the degree of selection on observables and an assumption on how much of the residual variation in the outcome is explained by the unobservables. There is strong evidence that my employment result is robust to selection on unobservable pre-migration characteristics, and moderate evidence that the earnings effect is robust. These results give me a degree of confidence that the true causal effect of workplace composition is non-null, and is plausibly close to the estimated associational effect.

Third, I allow for the possibility that my model is somehow misspecified. This is a relevant possibility, since the small sample I work with does not allow me fully flexibly incorporate my proposed controls in all estimation specifications without potentially overfitting the data. I therefore use variable selection methods for treatment effects in the presence of high-dimensional controls (Belloni et al., 2012, 2014; Chernozhukov et al., 2015), to semi-parametrically estimate the effect of the initial conational share on subsequent outcomes. The estimates are semi-parametric in the sense that I retain the

[1]This is assuming the bias from selection into return migration is not so large as to change the sign of the estimated effect.

assumption of linearity for the effect of the conational share, but allow the covariates to flexibly enter the regression function. The employment effects are highly robust, and become significantly negative sooner, after three years, while the earnings effects are less robust, in some cases disappearing.

Finally, I review the evidence for different mechanism 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. Instead, I suggest that a higher initial conational share is 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 second contribution of this paper is to instead focus on the initial place of work and composition of the set of coworkers. The switch of focus is novel, and 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 own group composition, 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 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 paper 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 first job on an immigrant's later labour market outcomes, to structure the empirical analysis. In Section 3 I discuss the data used for this project. In Section 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 5 I discuss different forms of selection, selection into return migration on the initial conational share and selection into the initial conational share on pre-migration characteristics and the bias they might impart on my estimates, and additionally explore semi-parametric estimates of my effect. In Section 6 I assess different possible explanations of my results. Finally, section 7 concludes and discusses avenues for future research.

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.

2.1 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.[2] 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 (1999) 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 misunderstand-

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

ings in the workplace, lowering productivity. Empirical evidence 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 (1999) 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.[3]

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 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.[4]

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. If firms learn about workers' productivity on the job, then the use of referrals also implies that turnover 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, the wage effect of the conational share will fade with tenure, since workers

[3] 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.

[4] 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.

hired in low conational-share firms, i.e. without a referral, will stay if the firm receives good news about their productivity, and adjusts the wage accordingly, while the wage effect of compensating differentials will last for the duration of the job.

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 have affects 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-document tendency towards homophily in the constitution of social networks (Mcpherson et al., 2001). In the terminology of Granovetter (1995), conationals would 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.[5]

The initial conational share might also affect subsequent outcomes 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 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

[5] 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.

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 unemployment, 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 term. This is particularly likely to be true if one accounts for the method of finding the first job, removing the negative effect of a referral on job separations.

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 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. 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 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 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 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.

4 OLS analysis

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}.

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

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 gender, 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 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 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 3 that the included pre-migration 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}.[6]

While the results presented in Table 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 structure of Oster's test in greater detail in Section 5.[7]

[6]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).

[7]In earlier versions of this paper, I have also considered instrumental variables estimates of the effect of the conational share on subsequent outcomes. Asylum seekers and ethnic Germans emigrating from

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 set of fixed effects 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 dummy 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. The full specification is therefore

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

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

4.2 OLS results

4.2.1 Employment rates

In Table 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), I first estimate the average effect of the initial conational share over with subsequent employment, first without controls (column 1), then with controls, including the other immigrant share (column 2). I find

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 my sample (around 200 individuals). The instrument is not strong enough to predict the conational share in such a small sample.

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 gender, 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 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), 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 5.1.

In the top panel of Figure 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 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.

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 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 5 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.

4.2.2 Wage earnings

In Table 6 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 hors worked, I repeat the estimation for respectively full- and part-time workers in columns 2 and 3. Part-time status and the daily wage are here measured on June 30 of a given year. 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 6 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 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 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 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, 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 bias that could explain this finding.

5 Potential sources of bias

5.1 Selection on unobservables into the treatment

The central identification claim of this paper 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-migration 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.

5.1.1 Overview

Intuitively, the test that I will apply involves comparing two sets of 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 variables on which selection takes place. Because selection is thought to explain a part of the uncontrolled association, one expects the coefficient 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 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 $R_{max} = \min\{1.3 \times \tilde{R}, 1\}$, where $\tilde{R}$ 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).[8]

5.1.2 Results

Table 8 reports the estimated values of Oster's δ for my employment specification in column 1. In the static specifications, in the first row, the 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 delta 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.

In columns 2 and 3 I report estimates of δ for the effect of the conational 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 including covariates increases the estimated effect and 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,

[8]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.

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.[9]

The results of these tests for selection on unobservables strengthens the claim that the associational effect of the initial coworker share on subsequent employment, estimated in Section 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.

5.2 Selection on the treatment into return migration

Having formally established that selection on unobservables is unlikely to explain the effects estimated in Section 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.

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

[9] 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.

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:

$$Y = a + \beta S + \varepsilon_Y. \tag{3}$$

The structural error term ε_Y is mean-zero[10] and independent of S, since there is no confounding. To model selection, I assume that latent utility C^* is a linear function of S, Y, and a mean-zero structural error term:

$$C^* = \alpha_S S + \alpha_Y Y + \varepsilon_{C^*}, \tag{4}$$

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:

$$C(S, Y) = \begin{cases} 1 & \text{if } C^* < K, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Equation (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, the true parameter of interest, β, can be defined as

$$\beta = \frac{\text{Cov}(Y, S)}{\text{Var}(S)} \tag{6}$$

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

$$\begin{aligned} \hat{\beta} &= \frac{\text{Cov}(S, Y | C = 0)}{\text{Var}(S | C = 0)} \\ &= \beta + \frac{\text{Cov}(S, \varepsilon_Y | C = 0)}{\text{Var}(S | C = 0)} \\ &= \beta + \frac{\text{Cov}(S, \varepsilon_Y | C^* \geq K)}{\text{Var}(S | C^* \geq K)} \end{aligned} \tag{7}$$

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 $\text{Cov}(S, \varepsilon_Y) = 0$ in the full sample by assumption, but not in the restricted sample of non-return migrants. The sign of the conditional covariance can

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

be calculated as

$$
\begin{aligned}
&\text{Cov}(S, \varepsilon_Y | C^* \geq K) \\
&\quad = E[S\varepsilon_Y | C^* \geq K] - E[S | C^* \geq K] E[\varepsilon_Y | C^* \geq K] \\
&\quad = E[\varepsilon_Y | C^* \geq K, S = 1] \Pr(S = 1 | C^* \geq K) - E[S | C^* \geq K] E[\varepsilon_Y | C^* \geq K] \\
&\quad = \{E[\varepsilon_Y | C^* \geq K, S = 1] - E[\varepsilon_Y | C^* \geq K]\} \Pr(S = 1 | C^* \geq K),
\end{aligned} \tag{8}
$$

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 (8), $E[\varepsilon_Y|\cdot]$. Note, however, that ε_Y is a mean-zero random variable and that its distribution is truncated when calculating the expectations $E[\varepsilon_Y|\cdot]$. 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

$$
\alpha_Y \varepsilon_Y \geq K - (\alpha_S + \alpha_Y \beta) S - \alpha_Y a - \varepsilon_{C^*}, \tag{9}
$$

This inequality makes clear how the sign of the bias of $\hat{\beta}$ 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 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 $\hat{\beta} < 0$, while if they are of opposite signs, then $\hat{\beta} > 0$. Therefore, if the estimated $\hat{\beta} < 0$ 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 $\beta < 0$.

5.2.2 The sign of the bias in the presence of confounding

In Appendix 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 nonlinearly on more parameters, obscuring the nature of the selection problem created by conditioning the analysis on the endogenous variable C, which be more clearly shown using a causal graph. Figure 5 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 5, the causal effect $S \rightarrow Y$ 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 $S \leftarrow X \rightarrow Y$. 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 $S \rightarrow Y$.[11].

Focussing on selection into return migration, in the top graph of Figure 5, 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.

[11]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 A.3

5.2.3 Evidence of selection on the treatment into return migration

In Section 4.2 I estimated that $\hat{\beta} < 0$. 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.[12]

In Figure 6 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 7 are significant at the ten per cent level. However, the causal graph presented in Figure 5 makes clear that the unconditional association between S and C plotted in Figure 6 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 7. The time trend is now statistically insignificant although the magnitudes are still economically relevant: an individual in her first year of employment

[12] 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.

in 2014 had a coworker share 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 $\beta < \hat{\beta}$.

5.3 Model misspecification bias

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 nevertheless 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 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 either 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).[13]

[13] Note the conceptual similarity of this approach to the Frisch-Waugh-Lovell theorem, where one

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.[14] 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 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

regresses the residual of a regression of Y on a low-dimensional X on the residual from regressing T on X.

[14] 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 is virtually unchanged, however interpreting the set of retained covariates is less straightforward.

second-largest group in my sample, was not necessarily expected *a priori*.

In the bottom panel of Figure 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 top 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 the bottom panel of Figure 4. These show that the modest 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 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 in Table 9.

To summarise, in this section I have shown that the employment effect estimated in Section 4.2.1 is robust to selection on unobservables, selective return migration, and more flexible regression specifications. There does not appear to be strong evidence that the earnings effect is different from zero.

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. 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 10 where I regress alternative outcomes on the conational share, conditional on the full set of controls and fixed effects defined in Equation (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 1 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 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.[15] 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 is a mechanism by which the initial conational share lowered employment in the longer term, it cannot 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 10 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 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

[15]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.

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 directly observing the use of networks or job search methods to find later jobs, I cannot positively show that these explain my finding, 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 the my finding. In columns one and two of Table 11 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 woman are typically less attached to the labour force than immigrant men, and have lower employment rates (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 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, the effect is also present for the other immigrant share. This is 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.

7 Conclusion

In this paper 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 suffer from selection bias due to differential selection into return migration based on the treatment of interest. This paper 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 paper could productively be 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 whether the effect is explained by the employment rate of one's past coworkers.

References

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

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

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

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

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

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

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

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

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

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

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

Battisti, M., G. Peri, and A. Romiti (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, 128–161.

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

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

——— (2014): “Inference on Treatment Effects after Selection among High-Dimensional Controls,” *Review of Economic Studies*, 81, 608–650.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gee, L. K., J. J. Jones, C. J. Fariss, M. Burke, and J. H. Fowler (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.

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

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

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

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

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

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

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

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

Kahane, L., N. Longley, and R. Simmons (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, 302–314.

Lazear, E. P. (1999): "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, 15–40.

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

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

Mcpherson, M., L. Smith-Lovin, and J. M. Cook (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, 1408–1418.

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

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

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

OECD (2020): "International Migration Statistics," .

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

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

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

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

Rosen, S. (1986): "The theory of equalising differences," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and R. Layard, Amsterdam: Elsevier, vol. I, 641–692.

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, 665–688.

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

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

8 Figures

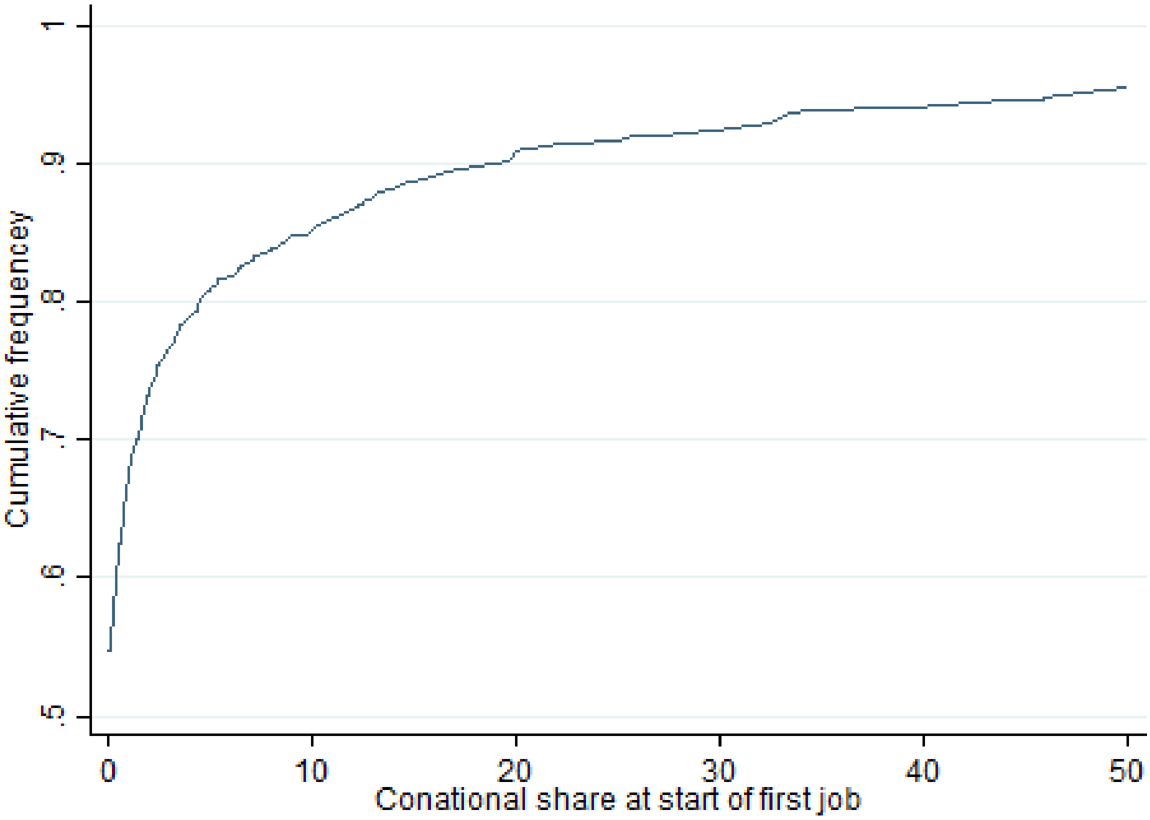

Figure 1: Empirical CDF of the initial conational share in the first job held by an immigrant in my sample. The distribution is truncated at 50, for ease of representation.

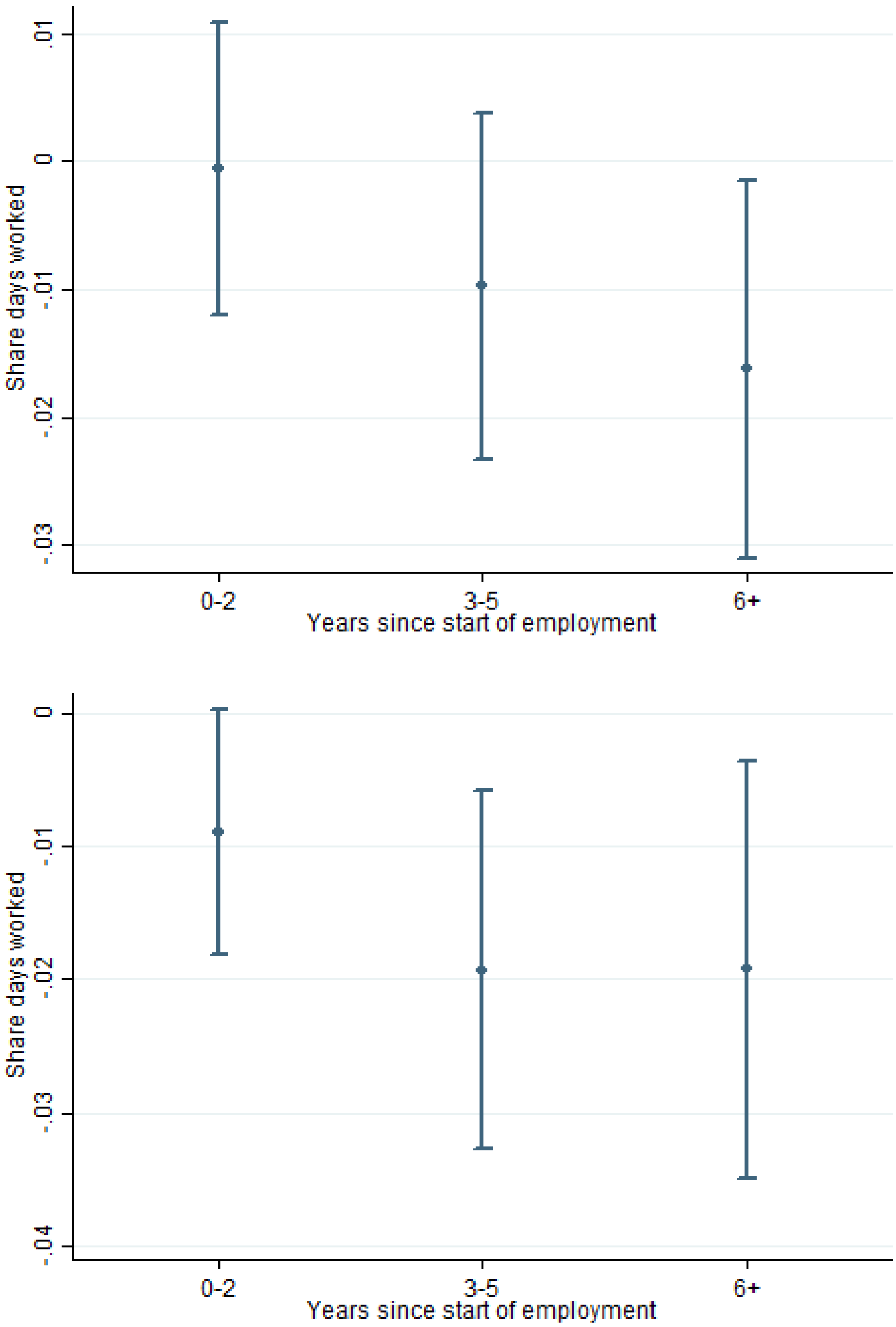

Figure 2: Dynamic estimates of the employment effect of the initial conational share. The top panel reports the coefficients from OLS estimates, the bottom panel reports semi-parametric estimates using the post-regularisation method of (Chernozhukov et al., 2015). 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

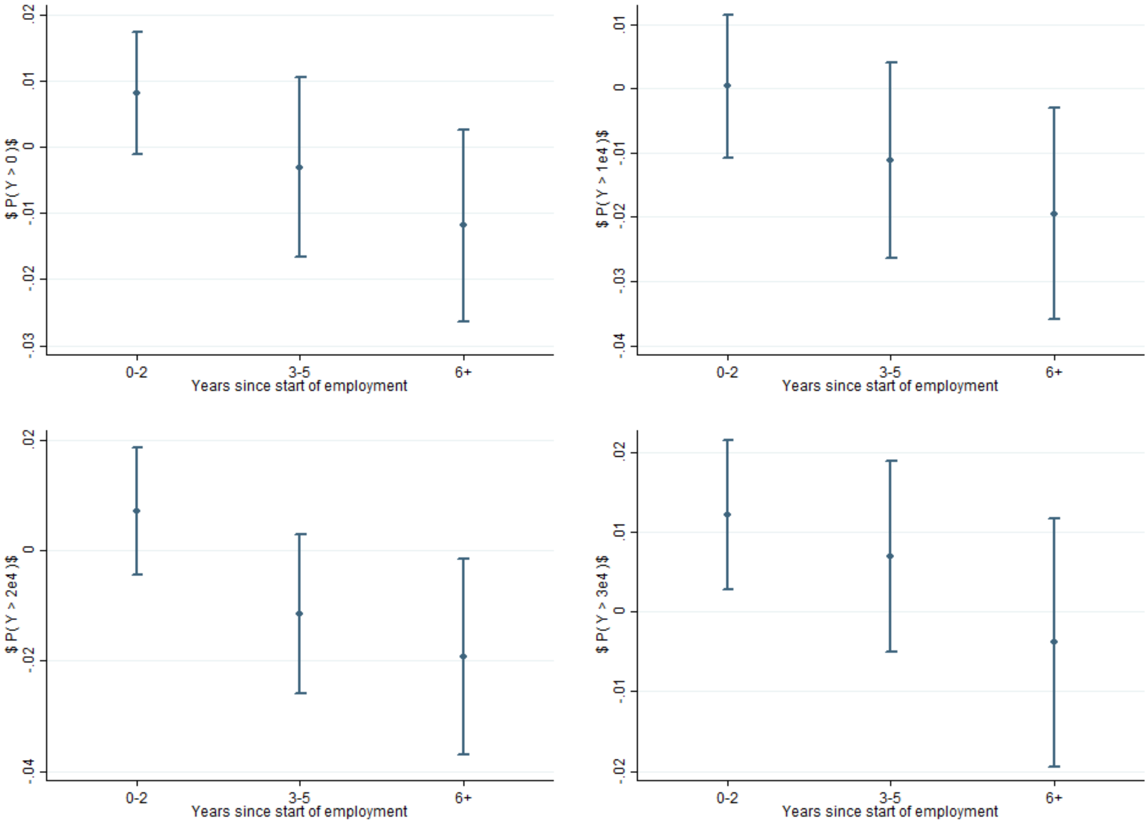

Figure 3: OLS estimates of the dynamic earnings effect of the initial conational share for different thresholds. 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

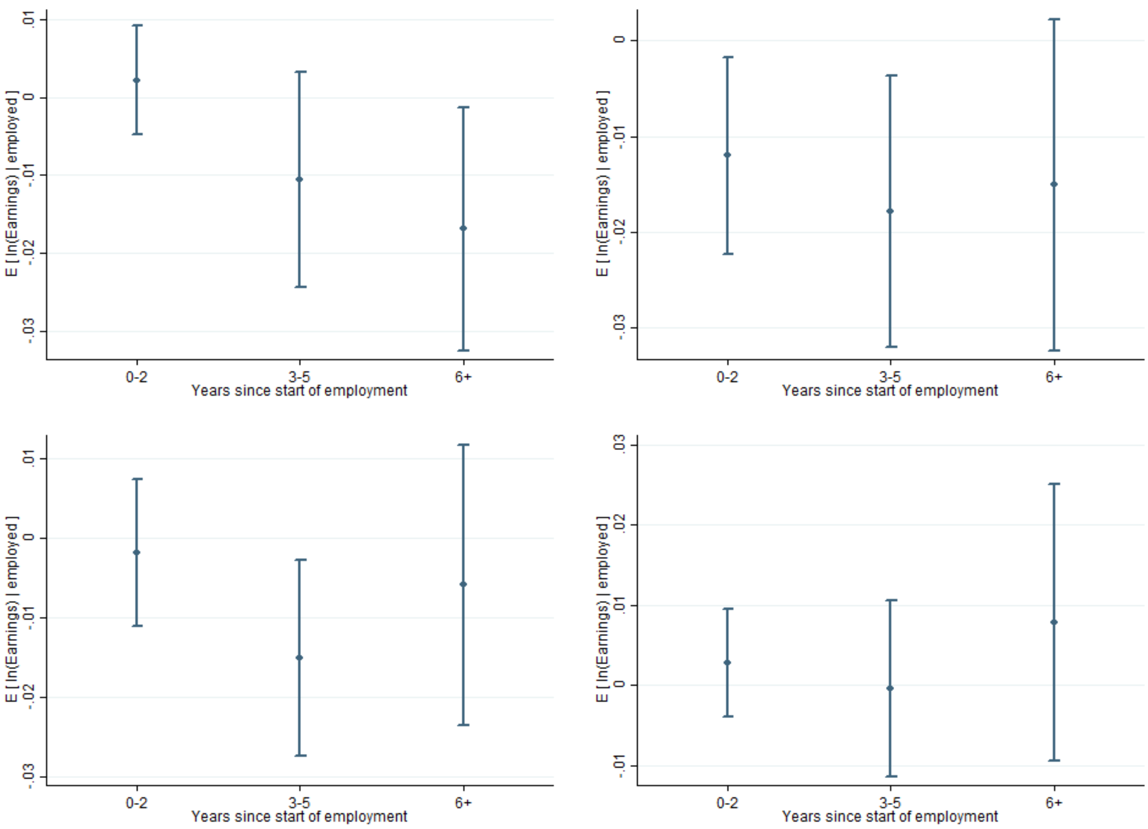

Figure 4: Post-regularisation estimates of the dynamic earnings effect of the initial conational share for different thresholds. 95 per cent confidence intervals reported are calculated using standard errors clustered by individual.

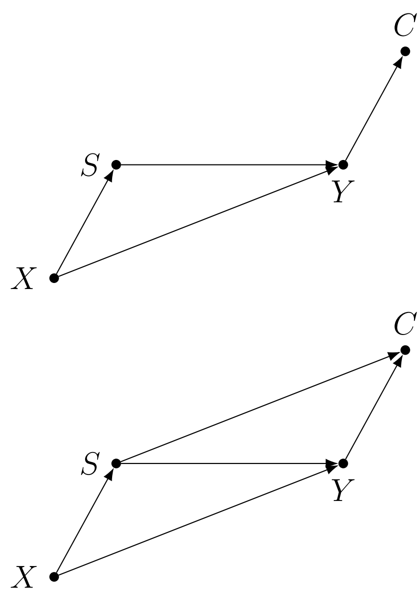

Figure 5: Possible causal structures relating initial coworker share S to subsequent employment Y, their common (observed) causes X, and return migration C. In the bottom panel, C is a collider along the path $S \to C \leftarrow Y$.

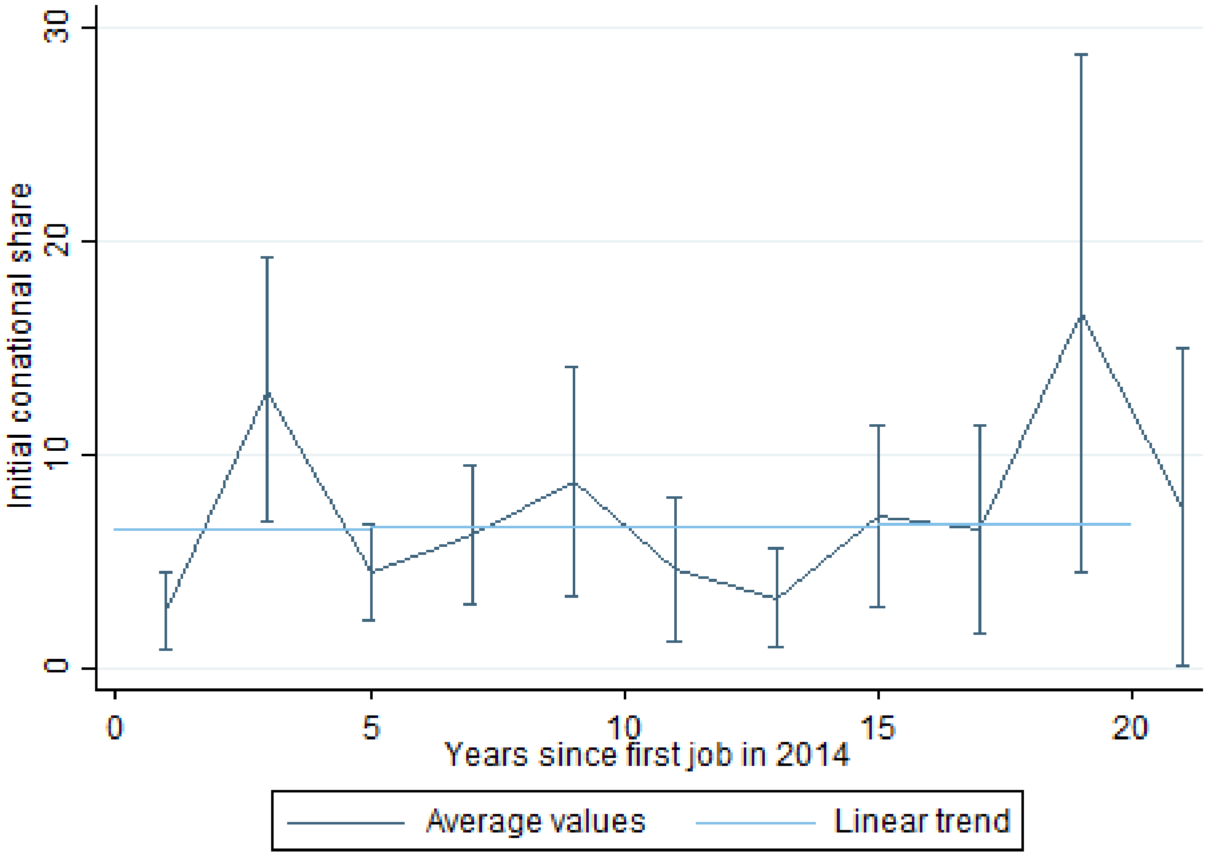

Figure 6: Differences in initial conational share for immigrants who first worked in Germany t years ago in 2014. Average values are grouped in two-year bins for data protection reasons.

9 Tables

Table 1: Summary statistics

	Mean	St. dev.	N
Panel A			
Share days employed	0.74	0.38	10068
Annual wage earnings	21250.2	15023.8	7497
P(Earnings > 10,000)	0.57	0.49	10068
P(Earnings > 20,000)	0.37	0.48	10068
P(Earnings > 30,000)	0.18	0.39	10068
0-2 years since start	0.25	0.44	10068
3-5 years since start	0.23	0.42	10068
6+ years since start	0.52	0.50	10068
Panel B			
Female	0.51	0.50	846
Low education	0.40	0.49	846
Medium education	0.31	0.46	846
High education	0.28	0.45	846
Age at migration	29.0	8.64	846
Employed before migrating	0.70	0.46	846
Panel C			
First job through contacts	0.57	0.50	846
Years until first job	3.28	3.02	846
Daily wage	43.1	34.4	846
Firm size	440.2	2082.6	846
Firm median wage	74.3	39.6	846
Firm age	13.1	10.5	846
Conational share	7.14	19.6	846
Other migrant share	16.6	20.4	846

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: Country groups

	N	Share
Russia	311	36.76
Romania	114	13.48
Poland	92	10.87
ex-Yugoslavia	70	8.27
Turkey	64	7.57
Asia	52	6.15
Italy	40	4.73
Other Europe	37	4.37
Africa	29	3.43
Greece	22	2.60
Others	15	1.77
Total	846	100.00

Note: Refers to country of birth for individuals born without German nationality.

Table 3: Association with other firm/job characteristics

	(1)	(2)
	$\beta^{s_i^{own}}$	$\beta^{s_i^{own}}$
Job characteristics		
log(Wage)	-0.02	0.00
	(0.02)	(0.02)
Apprentice	-0.00	-0.00
	(0.00)	(0.00)
Part-time	-0.01	-0.00
	(0.01)	(0.01)
Job through contacts	0.02^{*}	0.02^{+}
	(0.01)	(0.01)
Years until first job	0.05	0.04
	(0.06)	(0.06)
Firm characteristics		
log(Firm size)	-0.34^{**}	
	(0.02)	
log(Median wage)	-0.07^{**}	-0.03^{**}
	(0.01)	(0.01)
Firm age	-1.13^{**}	-0.39^{*}
	(0.16)	(0.17)
Firm age^2	-30.39^{**}	-6.56
	(4.92)	(5.17)
Other imm. share	-0.71^{**}	-0.17^{**}
	(0.03)	(0.04)
N	833	833

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

Table 4: Relation between initial coworkers and employment

	(1)	(2)	(3)	(4)	(5)	(6)
Conational share	-0.017**	-0.011+	-0.012**	-0.0043	-0.0050	-0.00057
	(0.0054)	(0.0055)	(0.0045)	(0.0052)	(0.0053)	(0.0059)
Conat. share × $\mathbf{1}(t \in [3,5])$			-0.0070	-0.0082	-0.0086	-0.0092
			(0.0064)	(0.0063)	(0.0063)	(0.0063)
Conat. share × $\mathbf{1}(t \geq 6)$			-0.0060	-0.012	-0.013	-0.016+
			(0.0089)	(0.0085)	(0.0084)	(0.0085)
Other mig. share		-0.0048				-0.0037
		(0.0049)				(0.0045)
Other mig. share × $\mathbf{1}(t \in [3,5])$						-0.0020
						(0.0048)
Other mig. share × $\mathbf{1}(t \geq 6)$						-0.0014
						(0.0062)
Premigration controls	No	Yes	No	No	Yes	Yes
Job controls	No	Yes	No	No	No	Yes
Firm 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	0.14	0.01	0.10	0.12	0.14
FE	No	Yes	No	Yes	Yes	Yes
Long-run coefficient			-0.0185	-0.0162	-0.0180	-0.0162
Long-run p-value			0.019	0.025	0.0100	0.031

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. + $p<0.1$, * $p<0.05$, ** $p<0.01$

Table 5: Relation between initial coworker share and other labour market outcomes

	(1) Conat. share	(2) Job separation	(3) Benefit receipt	(4) Jobseeker	(5) In training	(6) Not in IEB
Conational share	0.073**	0.011*	-0.0017	-0.0024	-0.0043**	-0.0045
	(0.0045)	(0.0049)	(0.0055)	(0.0059)	(0.0013)	(0.0034)
Conat. share × $\mathbf{1}(t \in [3,5])$	-0.034**	-0.0091	-0.000016	0.0010	0.000041	0.0085
	(0.0067)	(0.0064)	(0.0056)	(0.0056)	(0.0015)	(0.0054)
Conat. share × $\mathbf{1}(t \geq 6)$	-0.054**	-0.0098+	0.0072	0.0092	0.0019	0.017*
	(0.0064)	(0.0053)	(0.0074)	(0.0079)	(0.0014)	(0.0077)
Other mig. share	0.0011	0.0095*	-0.0032	-0.0097+	-0.0014	0.00090
	(0.0025)	(0.0042)	(0.0049)	(0.0050)	(0.0021)	(0.0024)
Other mig. share × $\mathbf{1}(t \in [3,5])$	-0.0034	-0.0048	-0.00025	0.0077	-0.00023	0.00086
	(0.0029)	(0.0060)	(0.0048)	(0.0051)	(0.0022)	(0.0031)
Other mig. share × $\mathbf{1}(t \geq 6)$	0.00083	-0.0027	0.014*	0.022**	0.0015	-0.0053
	(0.0042)	(0.0058)	(0.0069)	(0.0073)	(0.0027)	(0.0033)
Observations	7366	9911	9911	9911	9911	9911
Individuals	851	851	851	851	851	851
R^2	0.29	0.04	0.15	0.15	0.06	0.09
Sample						
Long-run coefficient	0.018	0.002	0.006	0.007	-0.002	0.013
Long-run p-value	0.002	0.683	0.437	0.357	0.077	0.067

Note: OLS estimates of relationship between initial conational share and other labour market outcomes. The conational share in later years (column (1)) is only defined when an individual is employed on June 30. 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. + $p<0.1$, * $p<0.05$, ** $p<0.01$

Table 6: Relation between initial coworkers and earnings

	(1) $\log(Y)$	(2) $P(Y>0)$	(3) $P(Y>1e4)$	(4) $P(Y>2e4)$	(5) $P(Y>3e4)$
Conational share	0.0020	0.0081^{+}	0.00041	0.0071	0.012*
	(0.011)	(0.0047)	(0.0057)	(0.0059)	(0.0048)
Conat. share $\times$ $\mathbf{1}(t \in [3,5])$	-0.012	-0.011^{+}	-0.012	-0.019**	-0.0053
	(0.015)	(0.0065)	(0.0078)	(0.0066)	(0.0052)
Conat. share $\times$ $\mathbf{1}(t \geq 6)$	-0.036^{+}	-0.020*	-0.020*	-0.026**	-0.016*
	(0.020)	(0.0078)	(0.0088)	(0.0089)	(0.0071)
Other mig. share	0.0028	-0.00030	0.0030	-0.0034	-0.0021
	(0.011)	(0.0035)	(0.0061)	(0.0052)	(0.0040)
Other mig. share $\times$ $\mathbf{1}(t \in [3,5])$	-0.0029	-0.0057	-0.0024	-0.0018	0.0021
	(0.016)	(0.0048)	(0.0069)	(0.0051)	(0.0037)
Other mig. share $\times$ $\mathbf{1}(t \geq 6)$	-0.013	-0.0037	-0.0092	0.0045	0.0037
	(0.016)	(0.0055)	(0.0078)	(0.0075)	(0.0070)
Observations	7366	9911	9911	9911	9911
Individuals	851	851	851	851	851
R^2	0.32	0.12	0.25	0.32	0.33
Long-run coefficient	-0.0336	-0.0118	-0.0194	-0.0192	-0.0038
Long-run p-value	0.0744	0.1093	0.0213	0.0342	0.6297

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 on June 30 of the relevant year. 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

Table 7: Relationship between initial conational share and return migration

	(1) Conat. share	(2) Conat. share	(3) Conat. share	(4) Conat. share
t	-0.49	2.66	5.12	3.33
	(1.11)	(8.76)	(9.42)	(10.6)
$t \times t$		-0.72	-1.14	0.031
		(1.99)	(2.10)	(3.03)
Average Employment rate			-5.85^{+}	11.5
			(3.28)	(20.0)
t × Average Employment rate				-6.43
				(7.07)
Constant	7.92**	4.90	5.26	1.41
	(3.02)	(8.82)	(8.92)	(6.36)
Observations	799	799	799	799
R^2	0.000	0.000	0.006	0.008

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 8: Estimates of Oster's δ

	$P(E=1)$	$P(Y>1e4)$	$P(Y>2e4)$	$P(Y>3e4)$	$P(E=1)$	$P(Y>1e4)$	$P(Y>2e4)$	$P(Y>3e4)$
Average effect	2.65	-11.2	-2.55	-1.02	1.62	-7.52	-1.69	-0.67
$t \in [0,2]$	1.41	-22.6	-2.14	-1.16	0.84	-15.0	-1.40	-0.75
$t \in [3,5]$	4.15	-69.8	-5.78	-2.35	2.49	-46.6	-3.79	-1.53
$t \geq 6$	6.20	-3.13	-1.73	-0.89	3.79	-2.10	-1.14	-0.58
Π	1.3	1.3	1.3	1.3	1.6	1.6	1.6	1.6
R^2_{max}	0.18	0.33	0.42	0.43	0.22	0.40	0.52	0.53

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 $R^2_{max} = \min\{\Pi\tilde{R}, 1\}$, where $\tilde{R}$ is the R^2 of the long regression, including all controls. For details, see main text.

Table 9: Semi-parametric estimates

	(1) $P(E=1)$	(2) $P(Y>0)$	(3) $P(Y>1e4)$	(4) $P(Y>2e4)$	(5) $P(Y>3e4)$
Conational share	-0.0089^{+}	0.0022	-0.012*	-0.0019	0.0028
	(0.0047)	(0.0036)	(0.0053)	(0.0047)	(0.0034)
Conat. share $\times$ $\mathbf{1}(t \in [3,5])$	-0.010	-0.013^{+}	-0.0058	-0.013*	-0.0032
	(0.0064)	(0.0067)	(0.0075)	(0.0064)	(0.0051)
Conat. share $\times$ $\mathbf{1}(t \geq 6)$	-0.010	-0.019*	-0.0030	-0.0041	0.0050
	(0.0089)	(0.0082)	(0.0087)	(0.0088)	(0.0083)
Observations	9911	9911	9911	9911	9911
Individuals	851	851	851	851	851
Medium-run coefficient	-0.019	-0.011	-0.018	-0.015	-0.00040
Medium-run p-value	0.005	0.132	0.014	0.016	0.943
Long-run coefficient	-0.019	-0.017	-0.015	-0.0060	0.0078
Long-run p-value	0.017	0.033	0.088	0.509	0.377

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

Table 10: Relation between initial coworkers and measures of social integration

	(1) Proficiency	(2) Naturalised	(3) Visited home	(4) Feel German	(5) Feel connected	(6) Foreign friends
Conational share	-0.047**	-0.0057	-0.047**	0.0067	0.021	0.032^{+}
	(0.017)	(0.0095)	(0.016)	(0.019)	(0.020)	(0.018)
Conat. share $\times$ $\mathbf{1}(t \in [3,5])$	0.011	-0.0062	0.022	-0.025	0.0072	-0.030
	(0.022)	(0.013)	(0.017)	(0.019)	(0.019)	(0.036)
Conat. share $\times$ $\mathbf{1}(t \geq 6)$	0.040*	-0.032**	0.056**	-0.024	-0.0079	-0.030
	(0.019)	(0.012)	(0.018)	(0.021)	(0.022)	(0.020)
Other mig. share	-0.042*	-0.036**	-0.049**	-0.021	-0.022	-0.023
	(0.019)	(0.012)	(0.017)	(0.015)	(0.019)	(0.027)
Other mig. share $\times$ $\mathbf{1}(t \in [3,5])$	0.019	0.018	0.030	0.010	0.022	0.047
	(0.022)	(0.018)	(0.021)	(0.021)	(0.023)	(0.031)
Other mig. share $\times$ $\mathbf{1}(t \geq 6)$	0.027	0.031*	0.052**	0.024	0.031	0.041
	(0.021)	(0.015)	(0.019)	(0.018)	(0.022)	(0.029)
Observations	1494	1483	1460	1477	1491	820
Individuals	837	835	837	832	837	820
R^2	0.29	0.39	0.28	0.15	0.15	0.16
Long-run coefficient	-0.007	-0.037	0.008	-0.017	0.013	0.002
Long-run p-value	0.472	0.000	0.311	0.059	0.186	0.880

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

Table 11: Heterogeneity of employment effect

	(1)	(2)	(3)	(4)	(5)
Conational share	-0.0070	0.0068	-0.0036	-0.0023	0.0077
	(0.0100)	(0.0074)	(0.0084)	(0.012)	(0.013)
Conat. share × $\mathbf{1}(t \in [3,5])$	-0.011	-0.012	-0.0015	-0.015	-0.012
	(0.010)	(0.0079)	(0.0078)	(0.012)	(0.018)
Conat. share × $\mathbf{1}(t \geq 6)$	-0.028*	-0.016	0.0015	-0.026	-0.039*
	(0.013)	(0.010)	(0.010)	(0.016)	(0.018)
Other mig. share	-0.013+	0.0061	-0.0086	-0.0023	-0.010
	(0.0069)	(0.0065)	(0.0081)	(0.0081)	(0.0081)
Other mig. share × $\mathbf{1}(t \in [3,5])$	-0.0015	-0.0040	0.0037	0.0065	-0.018+
	(0.0074)	(0.0061)	(0.0076)	(0.0072)	(0.010)
Other mig. share × $\mathbf{1}(t \geq 6)$	-0.0017	-0.0040	0.0035	0.0068	-0.024+
	(0.010)	(0.0073)	(0.010)	(0.0088)	(0.013)
Observations	4613	5298	4311	3160	2440
Individuals	428	423	338	270	243
R^2	0.17	0.20	0.18	0.22	0.26
Sample	women	men	low	med	high
Long-run coefficient	-0.035	-0.0088	-0.0022	-0.028	-0.031
Long-run p-value	0.007	0.275	0.792	0.030	0.056

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. 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). Columns (1) and (2) report results conditional on gender, columns (3)-(5) conditional on education level. Standard errors clustered by individual. + p<0.1, * p<0.05, ** p<0.01

A Alternative derivations of the sign of the bias induced by selective return migration

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 (8) 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

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

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 (8) 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

$$E[\varepsilon_Y | C^* \geq K, S = 1] > E[\varepsilon_Y | C^* \geq K] \tag{A.2}$$

and the bias will be positive.

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 5 is therefore

$$Y = \beta S + \Gamma_1 X + \varepsilon_Y \tag{A.3}$$

$$S = \Gamma_2 X + \varepsilon_S. \tag{A.4}$$

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 (A.3) and (A.4) 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:

$$C^* = \alpha_S S + \alpha_Y Y + \varepsilon_{C^*}, \tag{A.5}$$

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:

$$C(S, Y) = \begin{cases} 1 & \text{if } C^* < K, \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

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:[16]

$$\beta = E[\varepsilon_S^2]^{-1} E[Y \varepsilon_S] \tag{A.7}$$

However, conditioning on no return migration, $C = 0$, when estimating Equation (A.3) means that the OLS estimand is instead:

$$\hat{\beta} = E[\varepsilon_S^2 | C = 0]^{-1} E[Y \varepsilon_S | C = 0]. \tag{A.8}$$

[16] To verify Equation (A.7), substitute Equation (A.3) into Equation (A.7) and note that X and ε_S are uncorrelated by assumption.

Substituting Equation (A.3) into equation (A.9) and rearranging terms, one can show that

$$\begin{aligned}\hat{\beta} &= \beta + E[\varepsilon_S^2|C=0]^{-1}E[((\beta\Gamma_2+\Gamma_1)X+\varepsilon_Y)\varepsilon_S|C=0] \\ &= \beta + E[\varepsilon_S^2|C=0]^{-1}E[(Y-\varepsilon_S)\varepsilon_S|C=0]. \end{aligned} \tag{A.9}$$

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

$$\begin{aligned} E[(Y-\varepsilon_S)\varepsilon_S|C=0] &= \int yE[\varepsilon_S|C=0,y)dF_Y(y|C=0) - E(\varepsilon_S^2|C=0) \\ &= \int yE[\varepsilon_S|C*\geq K, Y=y)dF_Y(y|C^*\geq K) \\ &\quad - E(\varepsilon_S^2|C^*\geq K) \\ &= \int yE[\varepsilon_S|\alpha_S S+\alpha_Y y+\varepsilon_{C^*}\geq K)dF_Y(y|C^*\geq K) \\ &\quad - E(\varepsilon_S^2|C^*\geq K) \\ &= \int yE[\varepsilon_S|\alpha_S\varepsilon_S\geq K-\varepsilon_{C^*}-\alpha_Y y-\alpha_S\Gamma_2 X)dF_Y(y|C^*\geq K) \\ &\quad - E(\varepsilon_S^2|C^*\geq K). \end{aligned} \tag{A.10}$$

The second term of (A.10) 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. $\text{Cov}(S, C) < 0$, 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 $\hat{\beta}$ 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.

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:

$$Y = SY^1 + (1 - S)Y^0 \tag{A.11}$$
$$C = SC^1 + (1 - S)C^0. \tag{A.12}$$

I am interested in the causal effect of starting out in a high-conational share firm on the subsequent employment rate, $E[Y^1 - Y^0]$, 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 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:

$$E[Y|X, S = 1, C = 0] - E[Y|X, S = 0, C = 0]. \tag{A.13}$$

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

$$\begin{aligned}
&E[Y|X, S = 1, C = 0] - E[Y|X, S = 0, C = 0] \\
&\quad = E[Y^1|X, S = 1, C^1 = 0] - E[Y^0|X, S = 0, C^0 = 0] \\
&\quad = E[Y^1|X, C^1 = 0] - E[Y^0|X, C^0 = 0] \\
&\quad = \underbrace{E[Y^1 - Y^0|X, C^1 = 0]}_{\text{causal effect}} + \underbrace{E[Y^0|X, C^1 = 0] - E[Y^0|X, C^0 = 0]}_{\text{selection bias}},
\end{aligned} \tag{A.14}$$

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 $E[Y^0|X, C^1 = 0]$ and using the linearity of the expectations operator. Equation (A.14) 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, $E[Y^1 - Y^0|X, C^1 = 0]$, 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, $C^1 = 0$.[17]

The remaining terms of Equation (A.14) 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 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 $C^1 = C^0 = C$.[18] 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 $E[Y^1 - Y^0|X, C = 0]$, 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 (A.14) 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]In my parametric estimates in section 4.2 I further assume that the causal effect is constant over values of X, so the first term of equation (A.14) simplifies to $E[Y^1 - Y^0|C^1 = 0]$, an average treatment effect (ATE), rather than a conditional average treatment effect (CATE), i.e. an ATE conditional on X.

[18]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, $E[C^1 - C^0] = 0$. The stronger formulation here simplifies the exposition.