The workforce of pioneer plants

Ricardo Hausmann,* Frank Neffke†

April 11, 2016

Abstract

Is labor mobility important in technological diffusion? We address this question by asking how plants assemble their workforce if they are industry pioneers in a location. By definition, these plants cannot hire *local* workers with industry experience. Using German social-security data, we find that such plants recruit workers from related industries from more distant regions and local workers from less-related industries. We also show that pioneers leverage a low-cost advantage in unskilled labor to compete with plants that are located in areas where the industry is more prevalent. Finally, whereas research on German reunification has often focused on the effects of east-west migration, we show that the opposite migration facilitated the industrial diversification of eastern Germany by giving access to experienced workers from western Germany.

Acknowledgements

JEL codes

O33, R11, R12, O15, M13, M50, J23, J24

Keywords

diversification, technological diffusion, local labor markets, industry-specific human capital, mobility, clusters, Germany

---

*Harvard Kennedy School, Center for International Development and Santa Fe Institute
†Harvard Kennedy School, Center for International Development

Introduction

In the year 1838, the citizens of the Czech city of Plzen decided that enough was enough. After having been exposed to sub-standard quality for years, they took to the streets and spilled the beer of their local breweries into the city's rivers (DieWelt 2012). In response, the city council founded a new Burgher's Brewery. To ensure its quality, it recruited a young master brewer from Bavaria, Josef Groll. On October 5, 1842, Groll perfected his gift to the city that gave its name to the world's first Pilsner beer. Nowadays, Czech beer exports amount to well over 200 million USD and the Czech Republic is one of six countries in the world that report positive net beer exports to Germany.

Anecdotes like this one of how Plzen became famous for its beer through the mobility of skilled workers are widespread.[1] For example, in a rigorous exploration of the emergence of high-technology clusters in Taiwan, mainland China, India and Israel, Saxenian (2007) traces the origins of these clusters back to U.S. educated diasporas returning home, bringing with them the expertise and contacts that allowed them to start successful global companies. At the heart of these anecdotes and findings is that setting up new activities requires access to workers with the right skills and know-how. However, many important skills and much relevant know-how can only be acquired on the job through learning-by-doing. This means that firms that want to create new activities need access to workers with relevant work experience.[2] Indeed, over two thirds of all occupations listed in the US Department of Labor's occupational database O*NET require at least one year of relevant work experience. This poses a challenge for plants that belong to an industry hitherto absent from the region. By definition, for these plants, there are no local workers with industry experience. So how do such pioneer plants overcome this difficulty?

In this paper, we explore this predicament that emerges from the above-mentioned case study work more systematically. In particular, we ask how a new plant finds experienced workers if it is an industry's first local employer. How important is geographical mobility of workers in the spatial diffusion of industries? Are there differences among worker and industry types? And how much of this mobility was involved in the structural transformation of former East Germany after Germany's reunification?

To answer these questions, we start from a simple model in which plants face uncertainty about whether new hires without industry experience will meet the skill requirements of their jobs. However, this uncertainty is absent for experienced job applicants. Because pioneer plants can only hire experienced workers from outside the region, they have to balance the costs of non-local recruiting with the benefits of reducing risks. Whether it is optimal for a plant to hire local or non-local workers depends, among other things, on how strongly produc-

---

[1] Another example is the French Huguenots introducing wine in South African regions ((Bahar et al. 2014)).

[2] This number reflects the percentage of occupations in the O*NET 19.0 Database for which at least half of the respondents indicate that a minimum of one year of related work experience is required for the occupation.

tivity is affected when, inadvertently, a low quality worker is hired. Therefore, recruitment strategies should depend on the type of worker a plant is looking for. Moreover, recruitment strategies will also depend on whether plants can charge prices in excess of costs to absorb potential productivity shortfalls. Because pioneer plants face no competition from within the region, they have significant pricing power as long as their products are hard to trade across regions. This suggests that recruitment risks also depend on how tradable an industry's output is. Next, we extend the model to study why plants choose to locate in pioneer locations in the first place. We do this by assuming that plants require a mixture of workers with industry-specific skills (skilled workers) and without such skills (unskilled workers). This allows pioneers to offset the disadvantage of a lack of locally available experienced *skilled* workers by a cost-advantage in *unskilled* local workers.

To test the predictions of these models, we study the workforce that new plants in Germany hire within the first five years of their existence. For this purpose, we use data from the German social security records, which cover about 80% of the German labor market. Overall, we find that pioneering is rare. Moreover, the degree to which a plant is an industry pioneer in its location affects its recruitment strategy. Controlling for industry and region fixed effects, pioneers hire more workers from outside the region, who, typically, have industry experience. In line with our model, the difference between pioneers and new plants in locations where the industry is prevalent (cluster locations) in terms of the geographical distance over which they hire workers is larger if workers are skilled (i.e., college-educated). Indeed, controlling for industry and region fixed effects, the gap between pioneers and plants in cluster locations in the likelihood that workers are hired from outside the region starts widening from 2.5% until the 60th wage percentile to almost 12% in the 95th percentile. Moreover, this gap is also larger for plants operating in tradable (i.e., manufacturing) industries. However, in spite of their non-local recruitment strategies, pioneers still hire relatively many inexperienced workers. In general, hiring inexperienced workers correlates negatively with survival rates. However, this negative association only exists in non-traded industries, where pioneers' local monopoly position allows them to engage in risky hiring strategies. Moreover, we find that pioneers seem to leverage a low-cost advantage in unskilled labor by biasing their workforce composition toward the lower end of the skill spectrum. Finally, we explore to what extent the industrial transformation of eastern German regions depended on experienced labor from western Germany. Using a three-way interacted regression specification, we find that the pioneers that set up new industries in eastern Germany recruited substantially more of their experienced workers from western Germany than *vice versa*. The triple-differenced effects suggest that the excessive rate at which pioneer plants hire experienced workers from the other former territory is up to 38 percentage points (pp.) higher for college-educated workers in eastern German manufacturing pioneers than in their western German peers. Indeed, four out of five experienced workers in these eastern manufacturing pioneers were recruited from western Germany.

The fact that new plants hire key personnel from outside the region when lo-

3

cal experienced workers are unavailable suggests that the skills in the local labor market limit a region's scope for diversification. Given the low mobility of workers across country borders, this would explain why countries diversify slowly and incrementally (Hausmann and Klinger 2006). It also explains why multinationals are important in diffusing skills across the globe (Fosfuri et al. 2001, Keller 2004) and why these companies use expatriates to do so (Bonache and Brewster 2001, Hébert et al. 2005). Furthermore, our work is related to studies that attribute an important role to labor mobility in knowledge spillovers (Agrawal et al. 2006, Breschi and Lissoni 2009). Whereas these authors show how inventor mobility diffuses knowledge across regions, we show that the mobility of other types of labor helps diffuse industries. It also relates to the more general work on agglomeration externalities (e.g., Glaeser et al. 1992, Henderson et al. 1995, Rosenthal and Strange 2004), by empirically validating the importance of a local workforce with industry specific skills as hypothesized in this literature since the work of Marshall (1890). Finally, our work is also related to the literature on the home bias of entrepreneurs (Buenstorf and Klepper 2010, Dahl and Sorenson 2009, 2012), that with the exception of Timmermans (2010), has focused on the prior experience and whereabouts of entrepreneurs, not of a plant's first workers. We are unaware of any studies that show that new plants inside and outside clusters differ with respect to the geographic and other patterns of worker recruitment.

1   Industry pioneers and local labor markets

Economic convergence is a slow process at best. This holds as much for convergence among as within countries and is commonly ascribed to the fact that the superior technologies employed in advanced economies diffuse only slowly to lagging ones (Barro and Sala-i Martin 1992, Sala-i Martin 1996). Why however, in an age of almost costless and instantaneous communication, do industries and their technologies not diffuse more rapidly?

Tacit knowledge

A common response to this question in economic geography and international economics is that knowledge has a tacit component (Polanyi 1967). That is, for many of our capabilities, we are unable to explain verbally how they work. And even if we could, it does not follow that others would be easily able to replicate our actions. As a consequence, we can communicate only a small fraction of what we know and learning a skill often involves observation, repetition and coaching over prolonged periods of time in collaboration with people who possess the skill. Typical examples of this are learning how to ride a bike, how to play a musical instrument, how to speak a language or how to carry out a craft. However, tacit knowledge does not just matter in artisanal activities. It also plays an important role in high-technology, science-based industries (Collins and Pinch 1993, 1998). For instance, MacKenzie and Spinardi (1995) show that

4

apart from a deep understanding of nuclear physics, nuclear weapons design involves apprenticeship learning and expert (i.e., tacit-knowledge based) judgment. Access to tacit knowledge is therefore considered a *sine qua non* for firms to survive in almost any industry.

As a consequence of its tacit nature, important knowledge is often embedded in the experience and skills of workers. Therefore, new plants cannot simply buy or license a technology but also have to hire workers with the right skills and know-how. That explains why, in spite of high costs,[3] multi-national enterprises (MNEs) routinely reassign experienced employees to new branch plants abroad (Bonache and Brewster 2001). Interestingly, selection of expatriates is predominantly based on candidates' technical, not cross-cultural skills (Miller 1973, Mendenhall et al. 1987, Björkman and Gertsen 1993, Bonache and Brewster 2001) and expatriation is particularly successful in cases where industry-specific knowledge needs to be transferred (Hébert et al. 2005).

Labor market pooling and regional diversification

Not all plants have parent firms to provide them with an internal labor pool from which to recruit specialized workers. Instead, plants often rely on the local labor market to find skilled workers. Therefore, if skills are industry specific, the absence of other plants in one's industry constitutes a challenge. For instance, it is difficult to convince specialized workers to move to a region, unless there is a sufficiently thick market for their skills to mitigate hold-up problems. This coordination problem limits the geographical diffusion of economic activities and had already been recognized by Marshall (1890, IV.X.9), who pisted that: "[t]hese difficulties are still a great obstacle to the success of any business in which special skill is needed, but which is not in the neighbourhood of others like it." Marshall refers here to the problem of finding workers with specific skills in locations where these workers do not find much alternative employment and are therefore reluctant to move to. Interestingly, Marshall goes on to express his expectation that these problems were "... however being diminished by the railway, the printing-press and the telegraph." This findings in this paper, however, will suggest that Marshall's hope may have been too optimistic.

More recent work on agglomeration externalities puts forward that a specialized local labor force also generates strong externalities (Glaeser et al. 1992, Henderson et al. 1995) because thicker labor markets generate higher quality matches between employers and employees (Helsley and Strange 1990, Duranton and Puga 2004). The mobility of labor can give rise to further externalities through knowledge spillovers (Almeida and Kogut 1999, McCann and Simonen 2005). For instance, Agrawal et al. (2006) find that inventors are more likely to cite patents from their home regions. Similarly, Breschi and Lissoni (2009) find that inventors tend to cite work by former co-workers, suggesting that labor mobility not only diffuses existing knowledge across regions, but also leads to the creation of new knowledge.

---

[3]Peak (1997)reports that the typical expatriate costs over three times the employee's base salary.

Local labor markets also matter because they spawn the entrepreneurs needed to set up new plants. Often, these entrepreneurs start new firms in their home regions. This home bias may be due to personal reasons (Dahl and Sorenson 2009), but also to more tangible benefits. For instance, Dahl and Sorenson (2012) find that "region tenure," i.e., the number of years an entrepreneur has worked in the region, increases the survival chances of his or her firm by almost as much as industry tenure does. Dahl and Sorenson attribute this effect in part to entrepreneurs' know-who: entrepreneurs often draw on their local networks to find employees. Accordingly, local labor markets matter in helping firms gain access to the right employees.

However, if human capital is industry specific, such advantages are limited for pioneers. By definition, experienced workers cannot be hired on the local labor market if the plant is part of an industry that is new to the location. Consequently, pioneer plants, i.e., plants that are the first of an industry in their region, have two choices. Either they hire local workers without industry experience or they recruit experienced workers from elsewhere.

2   A model of recruitment by pioneer plants

To explore how the absence of experienced local workers affects the recruitment strategies of new plants, consider a model with two locations: a cluster location, where the industry is prevalent, and a pioneer location. We first focus on plants that are set up in pioneer locations and ask what determines whether they will hire local workers or recruit workers from other regions. The question of why pioneer locations are chosen in the first place is deferred to section 5.

In the model, plants produce one unit of output. The only input is labor, of which one effective unit is required. Furthermore, we assume that human capital is industry specific. Consequently, workers are not equally productive in each task. In particular, the amount of effective labor a worker provides in a given job depends on the extent to which her skills match the tasks the industry requires. Workers who are qualified for a job (H-type workers) produce one unit of effective labor, whereas unqualified workers (L-type workers) produce only q units of effective labor, where $0 < q < 1$. (Nominal) wages equalize within regions such that all local workers are paid the same wage, w.

We furthermore assume that, *a priori*, whether or not a worker is qualified for a particular job is unknown to both the employee and to the prospective employer. However, prior experience provides a reliable signal for industry-specific skills. That is, we assume that all workers with prior experience in an industry have proved their aptitude at the tasks this industry requires. In contrast, workers without industry experience dispose of the right skills only with probability $0 < \lambda < 1$. Finally, pioneers sell their unit of output at a price π, which, given that pioneers are the only local producers, will lie above the (perfectly competitive) price in clusters as long as goods are not perfectly tradeable.

When pioneers hire workers from the local labor market, by definition, these

workers will be inexperienced. In this local hiring strategy, H-type workers are hired with probability p and the expected costs of production for pioneers are:

$$pw + (1 - p)\frac{w}{q}$$

A pioneer can eliminate the uncertainty about a worker's skills by hiring experienced workers from outside the region. To attract these experienced workers, the pioneer must pay a relocation fee of R. This non-local hiring strategy yields the following costs:

$$w + R$$

Finally, if a plant fails to recover its costs, it incurs additional bankruptcy costs proportional to its losses.[4] Total costs under the local hiring scheme are therefore as follows:

$$C = pw + (1 - p)\frac{w}{q} + (1 - p)\theta \max\left(0, \frac{w}{q} - \pi\right)$$

Note that, because the non-local hiring strategy does not involve any uncertainty, bankruptcy risk exists only for the local hiring strategy. The cost function for producing with (experienced) non-local workers is therefore simply:

$$C = w + R$$

Two questions arise. First, is production feasible in the pioneer location? That is, is there a situation in which π exceeds costs? Second, should the pioneer plant risk hiring inexperienced local labor or should it attract experienced workers from outside the region? That is, when do the costs of the non-local or outside hiring strategy exceed the expected costs of the local hiring strategy?

Production with outside workers is feasible as long as:

$$\pi > w + R$$

Pioneer production with local workers is feasible if:

$$\pi > pw + (1 - p)\left(\frac{w}{q} + \theta \max\left(0, \frac{w}{q} - \pi\right)\right)$$
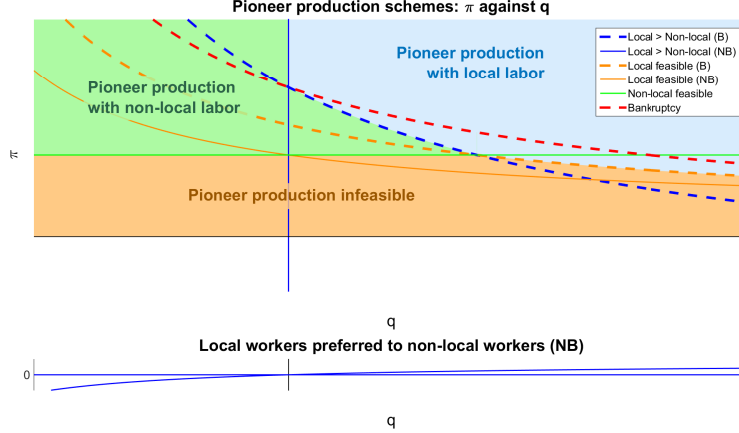
If $\pi \geq \frac{w}{q}$, local prices are so high that a plant avoids bankruptcy regardless of whether it draws an H- or an L-type worker. In this case, production with local workers is feasible as long as:

$$\pi > pw + (1 - p)\frac{w}{q} \tag{1}$$

If $\pi < \frac{w}{q}$ the pioneer plant will go bankrupt if it draws an L-type local worker. Production is now only profitable in expected terms (and therefore feasible) if $\pi > pw + (1 - p)\left(\frac{w}{q} + \theta\left(\frac{w}{q} - \pi\right)\right)$, or:

---

[4]Proportional bankruptcy costs are a simple way to induce some degree of risk avoidance.

Figure 1: The recruitment strategy of pioneer plants

Curves for regimes where bankruptcy can occur (B) are dotted, curves for regimes for which bankruptcy risk is absent (NB) are solid. Bankruptcy risk exists in the area below the dotted red line.

$$\pi > \frac{pw + (1-p)(1+\theta)\frac{w}{q}}{1 + (1-p)\theta} \tag{2}$$

Assuming production is feasible, pioneers prefer hiring (inexperienced) local workers to hiring (experienced) non-local workers as long as the associated expected costs are lower, or as long as $w + R > pw + (1-p)\left(\frac{w}{q} + \theta \max\left(0, \frac{w}{q} - \pi\right)\right)$, which simplifies to:

$$\begin{cases} \pi > \frac{1+\theta}{\theta}\frac{w}{q} - \frac{R}{\theta(1-p)} - \frac{w}{\theta}, & \text{if } \pi < \frac{w}{q} \\ q > \frac{(1-p)w}{(1-p)w+R} = \frac{1}{1+\frac{R}{(1-p)w}}, & \text{if } \pi \geq \frac{w}{q} \end{cases} \tag{3}$$

These production feasibility and indifference curves are summarized in Figure 1.

There are three different parameters in Figure 1. First, p determines how hard it is to find local workers with appropriate skills. This parameter will depend on the degree to which a plant is a pioneer in its region. Because local production is more often feasible and preferable at lower values of π and q, as p increases, all feasibility and indifference curves shift down and to the left:

> *Hypothesis 1*: The greater the extent to which a plant is a pioneer, the less likely it is that it will hire local workers.

Local prices (π) depend, among other things, on a pioneer's monopoly power. If a pioneer produces a non-traded product, by definition it will be a local monopolist. If, in contrast, the product is tradable across locations, the pioneer's

pricing power is eroded by competition from other locations. Figure 1 shows that if the local price is higher, more workers will be hired locally by the pioneer.

> *Hypothesis 2*: Pioneers that produce non-tradables will hire more
>   local workers than pioneers that produce tradables.

The final determinant of pioneers' recruitment strategies is q. q represents the productivity difference between L-type and H-type workers. We interpret this parameter as a measure of the specificity of the human capital an industry requires: the higher the specificity, the greater the productivity discount of workers who do not dispose of the right skill set. This suggests the following hypothesis:

> *Hypothesis 3*: The lower the specificity of the required human capi-
>   tal, the more often pioneers will hire local workers.

3 Data

The HES data set

We test these predictions on the German Historic Employment and Establishment Statistics (HES) database.[5] The HES is based on Germany's social security records and contains information on daily wages, a range of sociodemographic variables (such as educational attainment, gender, and age) and the industry, occupation, and location in which an individual is employed. The data are highly reliable, since they are used to determine social security contributions. However, wages are censored at the contribution limit. Because wages are only of secondary importance in this paper, we simply impute missing values based on age, age-squared, educational attainment, broad occupational dummies and gender, following the approach described by Gartner et al. (2005) and then deflate them to 2005 prices using the consumer price index provided by the German statistical office. To appear in the HES, individuals must be subject to social security payments. Consequently, the data set excludes all those who are exempt from social security contributions, such as civil servants and self-employed individuals, which together constitute about 20% of the German labor force. Moreover, we exclude all individuals in training and in part-time jobs.[6] The resulting data set covers yearly records[7] for about 20 million workers in each year, who can be followed throughout their working lives from 1975 to 2010. However, because of changes in the industry classification system we limit our analyses to the period 1999-2008.[8]

---

[5]See Bender et al. (2000) for a detailed description of this database.

[6]When establishing the origins of a worker in terms of geography and industry, we use the full data set.

[7]The data refer to a worker's employment situation on June 30.

[8]The data are classified using the German *Klassifikation der Wirtschaftszweige* 2003 (WZ 2003) at the 5-digit level for 2002-2008. At the 4-digit level, this classification matches the European NACE 1.1 classification. In the years 1998-2002, industries are classified according

Identifying new plants

The HES establishment identifiers are somewhat peculiar. Hethey and Schmieder (2010, p. 12) explain that:

> "... the definition of an establishment in this system does not necessarily correspond to a meaningful economic unit like a firm or a plant. ... [Identification] numbers are allocated to each organizational unit in a specific region and industry consisting of at least one worker liable to social insurance."

Consequently, if a firm has multiple branches in the same industrial activity and district, these branches may have the same identifier. Moreover, there is no way to assess whether different branches belong to the same firm. We will proceed with these caveats in mind. Hethey and Schmieder also show that new establishment identifiers do not always reflect new establishments. Instead, firms may reorganize the legal forms of their establishments or establishments may simply be recoded for unknown reasons. To avoid that such issues impact on our analyses, we use Hethey and Schmieder's (2010) definition of new plants. In this definition, spurious recodings are identified by studying the chunkiness of inter-establishment labor flows. In particular, we only keep new establishments that are considered to be real entrants by Hethey and Schmieder, meaning that the share of workers originating from the same prior employer must be below 30%. Using these definitions and after dropping observations with missing values, there are about 760,000 new plants (about 46,000 of which in manufacturing industries) between 2000 and 2004, the hiring decisions of which we can follow up until 2008.

Defining the degree of pioneering

To determine the extent to which a new plant is a local pioneer, we calculate whether its industry is overrepresented in the region in the year before the plant enters the local economy. That is, we calculate the pre-entry location quotient of the local industry (i.e., region-industry cell) in which a new plant is founded:

$$LQ_{irt} = \frac{emp_{irt}/emp_{.rt}}{emp_{i.t}/emp_{..t}}$$

where, emp_{irt} is the employment in region r, industry i and year t and omitted indices indicate summation over the corresponding class. Values between

---

to the Klassifikation der Wirtschaftszweige 1993 (WZ 1993), which matches the NACE 1.0 classification at the 4-digit level. We harmonize industry codes in two steps. First, we use the fact that in 2003, all establishments carry classification codes in both systems. As long as an establishment does not change its WZ 1993 code as we move backward in time, we use the WZ 2003 code of the establishment in 2003 also for previous years. Next, we use the information in the year 2003 to construct a correspondence between the WZ 1993 and the WZ 2003 codes. This correspondence is unique for all but 59 WZ 1993 industries. From these 59 industries we construct 29 merged industries. In the second step, we use this correspondence to fill in any remaining gaps.

0 and 1 indicate that the region is underspecialized in the industry, whereas values between 1 and infinity signal overspecialization. Consequently, the distribution of LQ is strongly right-skewed. Because LQ often takes on a value of zero, we cannot resolve this by taking logarithms. Instead, we use the following transformation to obtain a more symmetrically distributed variable:

$$p_{irt} = -\frac{LQ_{irt} - 1}{LQ_{irt} + 1}$$

This transformation first maps values corresponding to underspecialization onto the interval -1 to 0 and values that indicate overspecialization onto the interval 0 to 1, with same degrees of under- and overspecialization equidistant from zero, and then multiplies the result by -1. The latter transforms the specialization index into a pioneering index that ranges from -1 (the industry is completely concentrated in the region) through 0 (the share of the industry in the region equals the national share) to 1 (there is no prior employment in the industry, the plant is a true pioneer).

p_{irt} is our main variable of interest. In principle, for true pioneer plants, where the pre-entry location quotient is zero, p_{irt} equals one. However, this definition leaves us with preciously few true pioneers. Moreover, true pioneering will mostly happen in small regions and in industries that are small in Germany as a whole. As a consequence, focusing on true pioneers would bias our analyses towards a highly non-representative group of regions and industries. We therefore use p to measure the *degree-of-pioneering*. The lower the pre-entry LQ of a plant's industry-region cell, the higher p, and the more of a pioneer the plant is. High pre-entry LQ values (low values of p) mean that the new plant is in an industry that is over-represented in that region. We will refer to plants in high pre-entry LQ (low p) locations as *cluster plants*.

Table 1 presents the average of p for all industry-region cells across all years, once unweighted and once weighted by the number of new plants created in the cell. On average for all cells regardless of whether new plants appear there or not, p is positive. This reflects the fact that most industries are located in only a few regions, which results in many industry-region cells from which industries are absent and where p equals 1. By contrast, when p is weighted by the number of new plants in a location-industry cell, the average drops substantially. This means that new plants are disproportionately set up in cells with low p values, i.e., new plants are predominantly created in regions with a prior presence of the industry. In other words, pioneering is rather rare.

4 Worker origins

To investigate where new plants recruit their workers, we first select all full-time employees hired in the first five years of a new plant's existence. For these workers, we determine in which industry and region they had worked in the previous year. Some new hires do not appear in the year before they enter the new plant. This may indicate that this is a worker's first job, but also that she

Table 1: Average degree of pioneering by sector

industries	unweighted	weighted by # new plants
all	0.39	0.01
manufacturing	0.52	0.07
non-manufacturing	0.32	0.01

came from abroad, was unemployed, self-employed, a civil servant, on maternity leave, taking additional training, *inter alia*. We will refer to these workers as "previously inactive". For all other workers, we characterize the transition from their old to their new job in terms of a geographical distance and an industry distance.

We measure these distances, both in a dichotomous way, using dummy variables to indicate whether a worker previously worked in a different industry ("industry switchers") or planning area ("region switcher"), and in a continuous way. The continuous geographical measure is the road distance in kilometers between the districts of the new and the old employer. The continuous industry distance measure is supposed to capture the skill distance between two industries. As in Greenstone et al. (2010) and Neffke and Henning (2013), we use inter-industry labor flows as an indication of similarity in industries' skill requirements.[9]

Let F_{ij} be the total number of workers who move from industry i to industry j.[10] Furthermore, $F_{i.}$ denotes the number of workers who move from industry i to any other industry, $F_{.j}$ the number of workers who move from any other industry to industry j and $F_{..}$ the total number of workers who change industries in a year. The skill relatedness from i to j is now defined as:

$$SR_{ij} = \frac{F_{ij}}{F_{i.}F_{.j}}F_{..}$$

Values between one and infinity indicate that the labor flow between two industries exceeds what one would expect had the flows been random, given the overall in- and outflows of workers in the industries. We call these industries *skill related*. Values between 0 and 1 indicate that industries are unrelated in terms of their skill requirements. Analogous to the transformation of LQ, we map values indicating a lack of relatedness onto the interval -1 to 0 and values indicating skill relatedness on the interval 0 to 1:

$$SR_{ij}^* = \frac{SR_{ij} - 1}{SR_{ij} + 1}$$

---

[9] The underlying idea is that, to avoid obsolescence of their human capital, workers move predominantly among industries that require similar skills.

[10] We exclude all job transitions that involve any of the newly founded plants from these flows.

12

Table 2: Worker origins

	mean	std. dev.
switching		
from inactivity	0.496	0.500
from other plants	0.504	0.500
in other industry	0.705	0.456
in other region	0.279	0.449
continuous distances		
road distance (km)	61.5	117.6
skill relatedness	0.349	0.607

In the remainder, we use the average of SR_{ij}^* across all years between 1999 and 2008 to measure the relatedness between a worker's old and new job.

Table 2 provides averages and standard deviations for each of these distances. Note that a sizeable group of workers was inactive in the year before being hired by a new plant. Apparently, new plants create jobs that allow a large number of workers to (re)enter the social security system. Because the distances defined above can only be determined for workers for whom the previous employer is known, we analyze the group for whom this is not the case separately. However, our main focus is on workers whose origins can be determined. Of this latter group, about 70% switch industries and somewhat over a quarter are hired from outside the region.

Industry and worker groupings

To capture the model's distinction between traded and non-traded goods, we split industries into manufacturing and non-manufacturing industries. Moreover, the distinction between skilled and unskilled workers, which will play a prominent role in the expansion of the model in section 5, is made based on whether or not a worker has a college degree. Workers with unknown education are kept in a separate category.

Origins of the workforce

For reasons explained in section 3, we consider pioneering as a matter of degree. Therefore, to assess the difference between pioneer plants and cluster plants, we study the correlation between the distance (either in a geographical or in an industry sense) over which a worker is hired and a plant's degree of pioneering. Positive correlations signal that pioneer plants hire workers over longer distances than cluster plants do. Likewise, negative correlations means that plants in clusters tend to hire their workers from farther away. To partial out confounding effects, we use a regression framework instead of reporting raw correlations. In particular, we estimate equations of the following form:

$$d_e = \alpha_t + \gamma_i + \delta_r + \beta_p p_{ir\tau-1} + \epsilon_e$$

α_t, γ_i and δ_r are, respectively, year of hiring, industry and region fixed effects. These effects absorb differences in mobility associated with general regional, industry and year characteristics, such as the remoteness of the region or the overall spatial clustering of an industry. The subscript $\tau-1$ in the degree-of-pioneering variable highlights that we measure the industry's concentration one year before the plant's entry year, τ. d_e is the distance between the old and the new job of an employee e who is hired in year t by a new plant in region r and industry i. The unit of observation is thus a worker. Regardless of whether d_e is dichotomous or continuous, we run Ordinary Least Squares (OLS) regressions with errors clustered at the new plant's region-industry cell. Furthermore, because so many new hires come out of inactivity, we also estimate this equation with d_e a dummy dependent variable that takes the value of one if the worker was not in the social security system in the year prior to her hiring.

Our model predicts that the difference in hiring decisions of pioneer plants and cluster plants will depend on worker and industry characteristics. To study how β_p differs across worker and industry types, we also estimate the following interacted equation:

$$\begin{aligned}
d_e = \alpha_t + \gamma_i + \delta_r + \beta_p p_{ir\tau-1} \\
+ \sum_{s\in\{C,NC\}} \beta_s s_e + \sum_{s\in\{C,NC\}} \beta_{p\times s} s_e p_{ir\tau-1} \\
+ \beta_T T_i + \beta_{p\times T} p_{ir\tau-1} T_i + \epsilon_e
\end{aligned} \tag{4}$$

where T_i is a dummy for traded industries (proxied as those that belong to the manufacturing sector) and s_e a dummy for workers' skill level (proxied by college degree, C, or no college degree, NC, with missing education codes as the omitted category).

Do pioneers hire more workers out of inactivity? Table 3 shows that this is indeed the case. A one-unit increase in degree of pioneering is associated with a 3.6 pp increase in the likelihood that a plant hires a worker who was previously inactive. To understand the magnitude of this effect, note that new plants are set up in locations with a degree-of-pioneering of, on average, around zero (see Table 1). Given that, for true pioneers, p equals 1, a one unit increase in p coincides roughly with comparing the average new plant to a true pioneer. Henceforth, we therefore refer to changes associated with a one-unit increase in p as the "effect of pioneering." When controlling for the worker's level of education, the point estimate on p drops slightly (to 3.4 pp.). However, we find no evidence that pioneers' tendency to recruit non-local workers increases with the level of education or the tradability of the industry.

The fact that pioneers hire more workers out of inactivity is in line with the spirit of the model in section 2, which posits that locating in pioneer locations makes it more challenging to hire experienced workers. In line with this, we find

14

Table 3: Pioneering and likelihood of hiring from inactivity

dep. var.: inactivity (y/n)	(1)	(2)	(3)
p	0.036***	0.034***	0.029***
	(0.004)	(0.004)	(0.004)
edu(C)		-0.110***	-0.110***
		(0.005)	(0.005)
edu(NC)		-0.078***	-0.078***
		(0.002)	(0.002)
p×edu(C)			0.006
			(0.011)
p×edu(NC)			0.006
			(0.006)
p×man			0.010
			(0.013)
constant	0.485***	0.528***	0.528***
	(0.008)	(0.007)	(0.007)
year dummies?	yes	yes	yes
industry dummies?	yes	yes	yes
region dummies?	yes	yes	yes
R-squared	0.035***	0.041***	0.041***
# obs.	5,100,761***	5,100,761***	5,100,761***

p-values: ***: .01, **: .05, *:.10, clustered standard errors at the industry-region cell in parentheses. The table reports outcomes of OLS regressions. Observations are workers in the year that they are hired within the first five years' of a new plant's existence. The dependent variable is a dichotomous variable that evaluates to 1 if in the year prior to being hired, the worker was not working in a job covered by the German social security system. edu(C) and edu(NC) represent workers with a college (*Fachhochschule* or *Universität*) degree, respectively without such a degree. The omitted category consists of workers with unknown education. man represents is a dummy variable for manufacturing industries. p is a plant's degree of pioneering. Industry dummies are defined at the 5-digit level, regions at the level of German spatial planning areas.

Table 4: Pioneering and likelihood of hiring from outside the industry

dep. var.: switch industry (y/n)	(1)	(2)	(3)
p	0.101***	0.100***	0.074***
	(0.005)	(0.005)	(0.005)
edu(C)		-0.067***	-0.066***
		(0.006)	(0.006)
edu(NC)		-0.055***	-0.055***
		(0.002)	(0.002)
p×edu(C)			0.028**
			(0.014)
p×edu(NC)			0.026***
			(0.007)
p×man			0.051***
			(0.016)
constant	0.709***	0.741***	0.740***
	(0.012)	(0.013)	(0.013)
year dummies?	yes	yes	yes
industry dummies?	yes	yes	yes
region dummies?	yes	yes	yes
R-squared	0.177***	0.180***	0.180***
# obs.	2,570,638***	2,570,638***	2,570,638***

Idem Table 3, but with the dependent variable representing whether a worker switched to the new plant from an industry other than the new plant's industry (i.e., was inexperienced), while dropping all workers who were hired from inactivity.

that, restricting the sample to workers whom we can link to a prior employer, pioneers tend to hire more workers who lack industry experience: a one-unit increase in p is associated with a 10.1 pp. increase in the likelihood that a worker had switched industries (Table 4). Moreover, pioneers also hire workers from less related industries (Table 5). The effect of pioneering is again substantial, corresponding to over a third of a standard deviation decrease in skill relatedness. Both phenomena are most pronounced in manufacturing industries, but are more or less the same for workers with and without college degrees.[11]

Apparently, pioneers respond to the absence of experienced workers in the local labor market by hiring more inexperienced workers. However, because our main focus lies on the diffusion of industries, we are particularly interested in the spatial component of recruitment strategies. Tables 6 and 7 show that pioneers also hire substantially more workers from outside the region. On average, being a pioneer is associated with a 4.9 pp increase in sourcing workers from other regions. Focusing on the interaction terms, this turns out to be even more

---

[11]However, workers with and without college degrees are more often hired from outside the industry than workers whose education is unknown. This effect is not visible in the regression based on skill relatedness.

Table 5: Pioneering and skill relatedness to the previous employer

dep. var.: skill relatedness	(1)	(2)	(3)
p	-0.126***	-0.123***	-0.120***
	(0.007)	(0.007)	(0.008)
edu(C)		0.176***	0.176***
		(0.007)	(0.007)
edu(NC)		0.099***	0.099***
		(0.003)	(0.003)
p×edu(C)			0.004
			(0.018)
p×edu(NC)			0.013
			(0.011)
p×man			-0.059***
			(0.021)
constant	0.360***	0.299***	0.300***
	(0.014)	(0.015)	(0.015)
year dummies?	yes	yes	yes
industry dummies?	yes	yes	yes
region dummies?	yes	yes	yes
R-squared	0.120***	0.127***	0.127***
# obs.	2,570,638***	2,570,638***	2,570,638***

Idem Table 4, but with the dependent variable representing the skill relatedness between the old and the new employer's industries.

Table 6: Pioneering and the likelihood of hiring from outside the region

dep. var.: switch region (y/n)	(1)	(2)	(3)
p	0.049***	0.049***	0.023***
	(0.005)	(0.005)	(0.006)
edu(C)		0.084***	0.087***
		(0.005)	(0.005)
edu(NC)		-0.009***	-0.008***
		(0.002)	(0.002)
p×edu(C)			0.079***
			(0.015)
p×edu(NC)			0.022***
			(0.008)
p×man			0.046***
			(0.012)
constant	0.258***	0.260***	0.258***
	(0.018)	(0.018)	(0.017)
year dummies?	yes	yes	yes
industry dummies?	yes	yes	yes
region dummies?	yes	yes	yes
R-squared	0.059***	0.061***	0.061***
# obs.	2,570,638***	2,570,638***	2,570,638***

Idem Table 4, but with the dependent variable representing whether or not a worker was previously employed in a different spatial planning region.

so in the traded industries of the manufacturing sector (where this increase is compounded by another 4.6 pp) and for college-educated workers (which adds 7.9 pp to the likelihood of recruiting workers from other regions). This means that in total, pioneers in traded industries that are hiring college-educated workers are 14.7 pp more likely to recruit these workers from outside the region than the average new plant. This represents a more than 50% increase over the average likelihood of hiring workers from other regions. Using log(road distance) as a dependent variable shows that for this group of workers the distance to their old job is 45% higher if they are hired by pioneers instead of the average new plant (see Table 7).

We have argued that hiring outside the region should help pioneers recruit experienced workers. This would mean that pioneers should hire in particular workers with relevant work experience from farther away. We explore this by regressing the road distance to the old job on the skill relatedness between a worker's previous and current industry.[12] That is, we run the following regression:
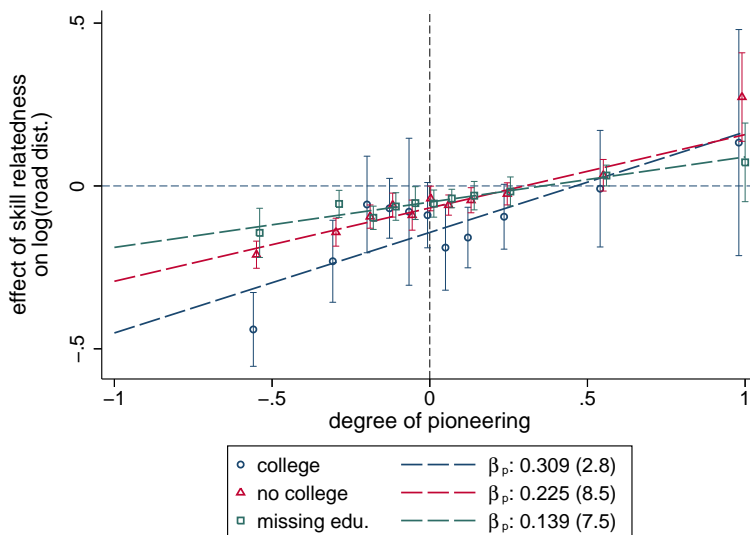
---

[12] Regressing region switching on industry switching yields similar results. However, because pioneers cannot hire local workers from within their industry, this regression is somewhat awkward, at least for the segment that only contains true pioneers.

Table 7: Pioneering and the road distance to the previous employer

dep. var.: log(road dist.)	(1)	(2)	(3)
p	0.090***	0.091***	0.014
	(0.014)	(0.014)	(0.016)
edu(C)		0.238***	0.245***
		(0.017)	(0.015)
edu(NC)		-0.015**	-0.014**
		(0.007)	(0.007)
p×edu(C)			0.218***
			(0.050)
p×edu(NC)			0.063***
			(0.021)
p×man			0.141***
			(0.029)
constant	3.321***	3.320***	3.317***
	(0.063)	(0.062)	(0.062)
year dummies?	yes	yes	yes
industry dummies?	yes	yes	yes
region dummies?	yes	yes	yes
R-squared	0.082***	0.084***	0.084***
# obs.	2,570,638***	2,570,638***	2,570,638***

Idem Table 4, but with the dependent variable representing the logarithm of the road distance to the previous employer's spatial planning region.

Figure 2: The effect of skill relatedness on road distance

The figure shows parameter estimates for different segments of the data of skill relatedness between a worker's old and new job in a regression of log(road distance) on skill relatedness, controlling for industry, year and region fixed effects. The degree-of-pioneering values represent the median value of a data segment.

$$\log\left(road\,dist_e\right) = \alpha_t + \gamma_i + \delta_r + \beta_{sr} SR_e^* + \epsilon_e$$

We repeat this regression for different segments of our data separately, where segments are created by grouping plants into 11 categories based on their degree of pioneering (with one category for true pioneers and the remaining categories taking more or less equal shares of workers in other plants). Figure 2 displays β_{sr} with its 95% confidence interval for each of these groups. Whereas for workers in plants with high degrees of pioneering, coming from a skill-related industry is associated with longer road distances, the opposite holds true for workers in cluster plants. This shows that pioneers tend to hire workers with relevant experience from farther away, whereas cluster plants attract *inexperienced* workers over long distances, while finding experienced workers close by. Moreover, cluster plants' advantage of being able to hire local workers with relevant work experience is most pronounced when it comes to hiring college-educated workers.

Survival rates of pioneers

Our model posits that pioneers run the risk of drawing L-type workers when hiring inexperienced workers. If pioneers act rationally, they should only be willing to accept this gamble and risk a costly bankruptcy if the scenario in

20

which they draw H-type workers yields positive profits. However, profits will be close to zero in traded industries, where cluster plants, who can hire experienced workers at no extra cost, drive profits down. Only in non-traded industries can pioneers balance the risk of negative profits with the positive profits granted by a local monopoly after a successful draw. Therefore, whereas we would expect some risk-taking, and thus, some failed entries, in non-traded industries, we would expect to observe much less of this in traded industries. That is, we would expect to see some additional plant failures in pioneer locations compared to clusters due to hiring inexperienced workers, but only in non-traded industries.

To test this, we compare how well the degree-of-pioneering predicts plant failure in non-traded and traded industries. In particular, we regress 5-year survival rates of new plants on their degree of pioneering, while splitting the sample into manufacturing and non-manufacturing plants. Next, we investigate to what extent the fact that pioneers hire more non-local workers can explain any association between plant survival and degree of pioneering in each sample. That is, we estimate the following model:

$$S5yr_p = \alpha_t + \gamma_i + \delta_r + \beta_p p_{ir\tau-1} + \beta_{isw} sh\_ind\_switch_p + X_p \psi + \epsilon_p$$

where $S5yr_p$ is a dummy that evaluates to 1 if plant p survives for at least 5 years and $sh\_ind\_switch_p$ denotes the share of first-year hires[13] who previously worked in a different industry and X_p is a set of variables that control for the quality of these first-year hires in terms of their educational attainment and whether or not they were previously inactive.

Results are reported in Table 8. In the overall sample, a one-unit increase in pioneering is associated with a 1.1 pp decrease in 5-year survival rates. That is, pioneers tend to have lower survival rates than cluster plants. However, we find this relationship only in the non-manufacturing sample. In manufacturing industries, pioneers do not suffer lower survival rates. Moreover, once we control for the share of workers in the plant who came from outside the industry, the effect of p becomes 14% less negative in the non-manufacturing sample.[14] Although this change is not statistically significant, it does show that whatever makes non-manufacturing pioneers struggle to survive correlates positively with the hiring of inexperienced workers.

5   Location choice

If pioneers experience difficulties in assembling an adequate workforce, why would they choose to locate outside clusters in the first place? Indeed, as long as wages equalize across locations, our model endows clusters with a cost advantage over pioneer locations, because they provide cheap access to experienced

---

[13] Controlling for workers that were hired in later years requires that the plant survives for those years, introducing undesired selection effects.

[14] These findings also hold when using average skill relatedness to old jobs of first-year hires and when adding a control variable for the plant's size at birth.

Table 8: Plant survival rates

dep. var.: 5 yr. surv. rate	(1) all	(2) all	(3) manuf.	(4) manuf.	(5) non-manuf.	(6) non-manuf.
p	-0.011***	-0.009***	0.005	0.005	-0.013***	-0.011***
	(0.002)	(0.002)	(0.006)	(0.006)	(0.003)	(0.003)
% inactive	-0.031***	-0.053***	-0.052***	-0.089***	-0.030***	-0.051***
	(0.002)	(0.003)	(0.009)	(0.017)	(0.002)	(0.003)
% college (C)	0.012**	0.013	-0.055***	-0.113***	0.016***	0.022**
	(0.005)	(0.010)	(0.019)	(0.040)	(0.005)	(0.011)
% no college (NC)	0.019***	0.028***	0.000	-0.040**	0.020***	0.032***
	(0.002)	(0.004)	(0.008)	(0.019)	(0.002)	(0.004)
% ind. switch		-0.030***		-0.048***		-0.030***
		(0.003)		(0.018)		(0.004)
ind. sw. × C		-0.002		0.076*		-0.008
		(0.012)		(0.046)		(0.012)
ind. sw. × NC		-0.016***		0.051**		-0.020***
		(0.004)		(0.021)		(0.004)
constant	0.443***	0.465***	0.569***	0.607***	0.433***	0.454***
	(0.010)	(0.010)	(0.023)	(0.027)	(0.011)	(0.011)
industry FE?	yes	yes	yes	yes	yes	yes
region FE?	yes	yes	yes	yes	yes	yes
year FE?	yes	yes	yes	yes	yes	yes
R-squared	0.077	0.077	0.037	0.037	0.079	0.080
# obs.	759,131	759,131	46,212	46,212	712,919	712,919

p-values: ***: .01, **: .05, *:.10, standard errors clustered at the region-industry level in parentheses. Dependent variable denotes whether or not the plant survived for at least 5 years.

22

workers. Consequently, pioneer locations have so far only been attractive if their local prices exceeded those in cluster locations. However, such price differences can only exist if output is costly to trade across locations. This would rule out pioneering in fully traded industries.

However, if pioneer locations had certain cost advantages of their own, these cost advantages could make them competitive alternatives for the production of tradable as well as non-tradable goods. We now explore what would happen if, to counterbalance the easy access to specific human capital in clusters, pioneer locations were to provide relatively cheap access to generic human capital. In other words, we investigate what happens if wages are not the same across locations. Note that, because we are interested in the wages that enter a plant's production function, what we require is that *nominal* wages differ. However, if costs of living can vary from place to place, this assumption does not preclude that *real* wages equalize across locations.

To model this, we assume that plants require a combination of skilled and unskilled labor to produce their unit of output. Unskilled workers have human capital that is generic (i.e., equally productive in all activities). Skilled workers, in contrast, have specific human capital. These skilled workers can, as before, be qualified (H-type), i.e., have suitable human capital for the industry in question, or not (L-type)[15] and prior industry experience is, once again, supposed to be a sufficiently strong signal to ensure that a skilled worker is of type H. For simplicity, we assume that unskilled and skilled workers are used in a fixed ratio, α. That is, production is Leontief, with no substitution between skilled and unskilled workers.

In what follows, we expand our notation with a locational superscript $l \in \{c, p\}$ – where c denotes the cluster location and p the pioneer location – and with a worker-type subscript $\sigma \in \{s, u\}$ – where s denotes skilled workers and u unskilled workers. The Δ symbol is used as an abbreviation for the differences between cluster and pioneer locations, that is $\Delta x \equiv x^c - x^p$.

Furthermore, we assume that pioneer locations have an unskilled-labor cost-advantage: $\Delta w_u \leq 0$. If they didn't, pioneer locations could only attract plants if their skilled labor were extraordinarily cheap ($\Delta w_s \gg 0$). However, this would mean that the cluster location has a comparative advantage in generic, not in specific skills, which goes against the spirit of our model.

Second, we exclude the possibility that the cluster attracts skilled workers from the pioneer location. This could happen if skilled workers in the pioneer location were so cheap that cluster plants would want to recruit them. However, because wages of unskilled workers are also assumed to be lower in the pioneer location, cluster plants should simply relocate to the pioneer location, making the cluster location unsustainable.

Third, we assume that relocation costs are sufficiently high to make hiring non-local unskilled workers unattractive for plants in clusters. That is, we assume that $R > \Delta w_u$.[16]

---

[15] Note that the L- and H-categories do not refer to the skill level, but to the match of skilled workers to their industries.

[16] Allowing $R > \Delta w_u$ does not substantially change outcomes, but means that the pioneer

23

Zero-profit assumption for cluster

We focus on the case for fully (i.e., costlessly) traded goods, the price level of which is anchored by the costs of production in clusters.[17] Inside clusters, competition is perfect so that cluster firms make zero profits. The only viable hiring strategy under these assumptions is that cluster plants produce with unskilled local and skilled local workers.

The costs in this scenario are:

$$C^c = \alpha w_u^c + w_s^c$$

Prices are set by the following zero-profits condition:

$$\pi^c - \alpha w_u^c - w_s^c = 0$$

For fully tradable products, this price carries over to all locations:

$$\pi^c = \pi^p = \pi = \alpha w_u^c + w_s^c$$

Pioneer production

With prices anchored by the costs in cluster locations, there are two potential hiring strategies for pioneers.[18] First, pioneers can hire both unskilled and skilled workers locally, yielding the following costs:

$$C^p = \alpha w_u^p + p w_s^p + (1-p)\left(\frac{w_s^p}{q} + \max\left(0, \theta\left(\alpha w_u^p + \frac{w_s^p}{q} - \alpha w_u^c - w_s^c\right)\right)\right)$$

Alternatively, pioneers could hire local unskilled workers and non-local skilled workers. Costs under this scenario are:

$$C^p = \alpha w_u^p + w_s^c + R$$

Location choice and hiring strategy now depend on α and q. Once again, we explore when pioneer production is feasible (i.e., under which circumstances pioneers have positive profits) and when local skilled workers are preferred to non-local skilled workers.

For fully tradable goods, pioneering is only feasible if production costs are lower in pioneer locations than in clusters:

$$\alpha w_u^c + w_s^c > \alpha w_u^p + p w_s^p + (1-p)\left(\frac{w_s^p}{q} + \max\left(0, \theta\left(\alpha w_u^p + \frac{w_s^p}{q} - \alpha w_u^c - w_s^c\right)\right)\right)$$

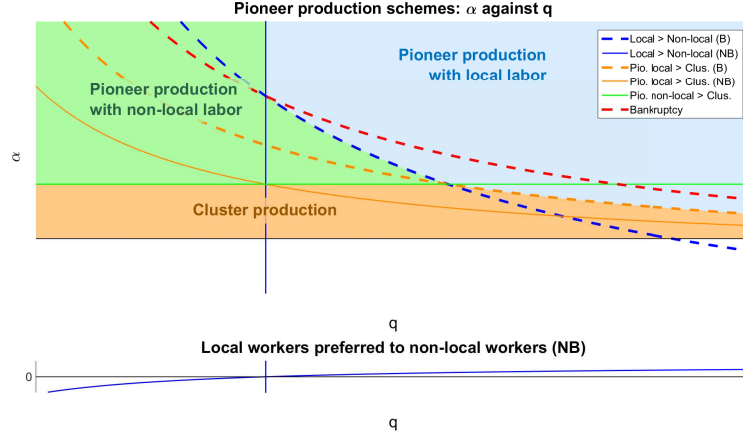This condition simplifies to:

---

cost advantage is not Δw_u, but $\max(R, \Delta w_u)$.

[17]The extension to partially traded goods is trivial, as it simply raises profitability in all pioneer production schemes uniformly by $\Delta\pi$.

[18]Note that unskilled labor will never be hired non-locally, because $\Delta w_u > 0$ by assumption.

Figure 3: Location choice

Idem Figure 1.

$$
\begin{cases}
\alpha \Delta w_u > \frac{p w_s^p + (1-p)(1+\theta)\frac{w_s^p}{q}}{1+(1-p)\theta} - (w_s^p + \Delta w_s) & \text{if } \alpha \Delta w_u < \frac{w_s^p}{q} - w_s^c \\
\alpha \Delta w_u > p w_s^p + (1-p)\frac{w_s^p}{q} - (w_s^p + \Delta w_s) & \text{if } \alpha \Delta w_u \geq \frac{w_s^p}{q} - w_s^c
\end{cases}
\tag{5}
$$

Pioneers prefer working with local skilled workers if:

$$
\alpha w_u^p + w_s^c + R > \alpha w_u^p + p w_s^p + (1-p)\left(\frac{w_s^p}{q} + \max\left(0, \theta\left(\alpha w_u^p + \frac{w_s^p}{q} - \alpha w_u^c - w_s^c \right) \right) \right)
$$

which simplifies to:

$$
\begin{cases}
\alpha \Delta w_u > \frac{1+\theta}{\theta}\frac{w_s^p}{q} - \frac{R+\Delta w_s}{(1-p)\theta} - \frac{w_s^p}{\theta} - (w_s^p + \Delta w_s) & \text{if } \alpha \Delta w_u < \frac{w_s^p}{q} - w_s^c \\
q > \frac{1}{\theta + \frac{\Delta w_s + R}{(1-p)w_s^p}} & \text{if } \alpha \Delta w_u \geq \frac{w_s^p}{q} - w_s^c
\end{cases}
\tag{6}
$$

Figure 3 depicts conditions (5) and (6) in a plot of α against q. Note that Figure 3 strongly resembles Figure 1. Indeed, a comparison of conditions (5) and (6) to their counterparts in the model of section 2 (conditions (1/2) and (3)) reveals that the feasibility and indifference curves have not substantively changed.[19] As a consequence, the hypotheses derived from the earlier model carry over to this location choice extension. Moreover, our assumption that effective wages are lower in pioneer locations yields a new hypothesis:

---

[19]Indeed, substituting π by $\alpha \Delta w_u$ in (1) and (2) yields a condition close to (5), but shifted by $-\left(w_s^p + \Delta w_s\right)$. Similarly, substituting $\alpha \Delta w_u$ for π yields a condition close to (3), but with the first parted shifted down by $-\left(w_s^p + \Delta w_s\right)$ and the vertical line in the second part shifted to the left by the Δw_s term in the denominator.

Hypothesis 4: Controlling for worker quality, wages are lower in pioneer locations than in cluster locations.

Finally, because pioneer production is more feasible and preferable to cluster production the larger is the required amount of unskilled workers, the location choice model suggests a fifth hypothesis:

Hypothesis 5: Pioneers use a lower share of skilled workers than plants in clusters do.

6   Location choice and workforce composition

Wages

To test hypothesis 4, we regress workers' wages on the degree to which their plant is a pioneer. To allow pioneers' wage advantages to differ by worker and industry type, we interact the plant's degree of pioneering with a worker's educational attainment and a dummy for manufacturing industries. Because we are interested in the costs of plants, we do not add region dummies to this regression and thus investigate "nominal" wages.[20] However, because we want to compare plants that operate on the same markets, we do control for industry fixed effects at the 5-digit level. Moreover, we add a vector of individual-level characteristics to control for differences in worker quality, which yields the following equation:[21]

$$\log\left(wage_e\right) = \alpha_t + \gamma_i + \sum_{s \in \{C,NC\}} \beta_s s_e + \beta_T T_i$$

$$+ \beta_p p_{ir\tau-1} + \sum_{s \in \{C,NC\}} \beta_{p \times s} s_e p_{ir\tau-1} + \beta_p^T T_i p_{ir\tau-1} + \epsilon_e$$

Table 9 shows that, even within the same industry, pioneers pay substantially lower wages than cluster plants do. The point estimate implies that, unconditionally, a one-unit increase in pioneering translates into 14% lower wages. Column 2 shows that although some of this difference is due to observable differences in worker quality and industry-specific effects, after accounting for such factors, a one-unit increase in pioneering still implies a wage advantage of 12%. Adding interactions of p with worker-type and industry-type dummies in column (3) shows that this wage gap is neither statistically different in manufacturing industries, nor for workers with a college degree.

Workforce composition

The fact that the association between wages and degree of pioneering weakens after controlling for worker characteristics suggests that pioneers hire lower-skilled workers. To explore whether pioneer plants indeed hire workers with a

---

[20]Among other things, region fixed effects would absorb differences in costs of living.

[21]Because we are interested in within-industry variation, we cluster standard errors by industry.

Table 9: Entry wages at new plants

dep. var.:	(1)	(2)	(3)
p	-0.154***	-0.124***	-0.150***
	(0.048)	(0.021)	(0.031)
edu(C)		0.917***	0.916***
		(0.020)	(0.020)
edu(NC)		0.332***	0.332***
		(0.017)	(0.017)
p × edu(C)			-0.010
			(0.042)
p × edu(NC)			0.036
			(0.022)
p × man			0.045
			(0.037)
male		0.376***	0.376***
		(0.018)	(0.018)
age		0.094***	0.094***
		(0.004)	(0.004)
age-squared		-0.001***	-0.001***
		(0.000)	(0.000)
foreign		0.011	0.011
		(0.016)	(0.016)
inactive		-0.368***	-0.368***
		(0.019)	(0.019)
constant	3.262***	1.539***	1.539***
	(0.057)	(0.086)	(0.086)
industry dummies?	no	yes	yes
region dummies?	no	no	no
year dummies?	yes	yes	yes
R-squared	0.002	0.341	0.341
# obs.	5,100,761	5,100,761	5,100,761
# clusters	919	919	919

p-values: ***: .01, **: .05, *:.10, clustered standard errors at the industry level in parentheses. The dependent variable is the natural logarithm of workers' starting wages in new plants. The worker characteristics we control for are: gender, age and age-squared, whether or not a worker is a foreign national and whether or not he or she was inactive in the year prior to being hired.

different skill profile, we calculate to what extent the occupational mix of a new plant's workforce resembles the one in their industry as a whole. Let o_a be a vector that contains the occupational employment shares for plant a of length O, with O the number of occupations in the economy. Similarly, let $o_{I(a)}$ be the vector of occupational employment shares in a's industry, $I(a)$. The closeness of a plant's occupational mix to the industry's average, i.e., its *skill fit*, can now be defined as the Pearson correlation between o_a and $o_{I(a)}$:

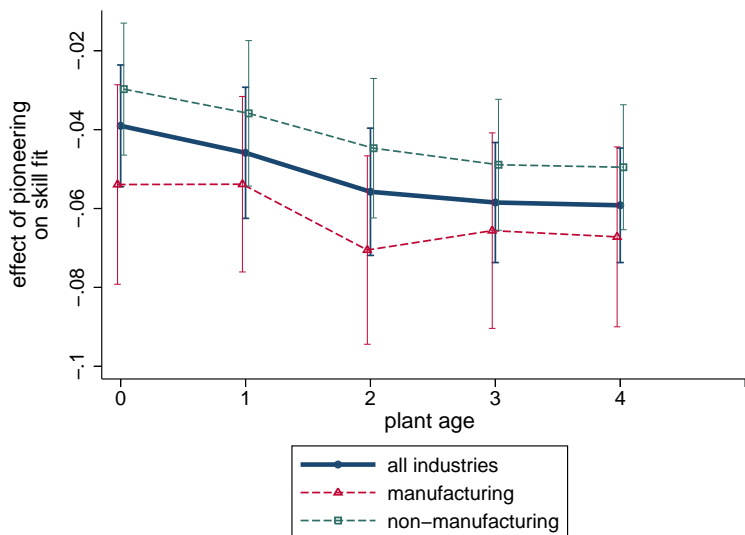$$skill\ fit_a = \text{Corr}\left(o_a, o_{I(a)}\right)$$

The average skill fit of new plants ranges from 0.50 at birth to 0.62 four years later. However, given the local availability of experienced workers, cluster plants should find it easier than pioneer plants to hire the right skill mix. Figure 4 shows the parameter estimate of a regression of a plant's skill fit on its degree of pioneering at ages 0 to 4.[22] Regardless of whether they are manufacturing or non-manufacturing plants, pioneers employ occupational mixes that resemble their industry's less than those of plants in cluster locations. Moreover, even though plants tend to gradually move towards the average occupational mix, the difference between cluster and pioneer plants grows with plant age. This suggests that, pioneers use a different technology from cluster plants.

Figures 5a to 5c show that pioneers indeed tend to hire lower skilled labor. Figure 5a shows the share of workers with and without a college degree as a percentage of all first-year hires, at different degrees of pioneering. To keep the comparison between plants that belong to the same industry, we show these shares in deviation of their industry-mean. Figure 5b shows the same graph, but now for the share of workers hired out of inactivity. Finally, Figure 5c shows the share of workers in non-ubiquitous occupations, i.e., occupations that are concentrated geographically.[23] Such rare occupations are likely to require highly specific skills. The results are remarkably unequivocal. Although differences amount to just a small number of percentage points, all three graphs confirm that even within narrowly defined industries, pioneers bias their workforce to lower-skilled workers and workers with less specific skills than cluster plants do.

---

[22]We control only for year effects. Adding controls for plant size does not make much difference. Adding industry fixed effects reduces the effect of the degree of pioneering by about 50%, but does not change the shape of the curve.

[23]To be precise, non-ubiquitous occupations are those with a high average squared difference between a region's share of nation-wide employment in the occupation and the region's share of total national employment: $U_o = \sum_r \left(\frac{emp_{or}}{emp_{o.}} - \frac{emp_{.r}}{emp_{..}} \right)^2$, where emp_{or} represents the number of workers in a region in an occupation and "." denotes summation over the omitted category. For occupations with a share of regional employment proportional to the region's overall size $U_o = 0$, whereas for occupations that are completely concentrated in one region U_o approaches 1. We repeat this for each year between 1999 and 2008 and then take averages. Occupations with below-median levels of U_o are considered ubiquitous, those above the median, non-ubiquitous.

Figure 4: Skill fit

The graph shows how the negative association of a plant's degree of pioneering with how similar its own occupational mix is to the industry's average changes with plant age.
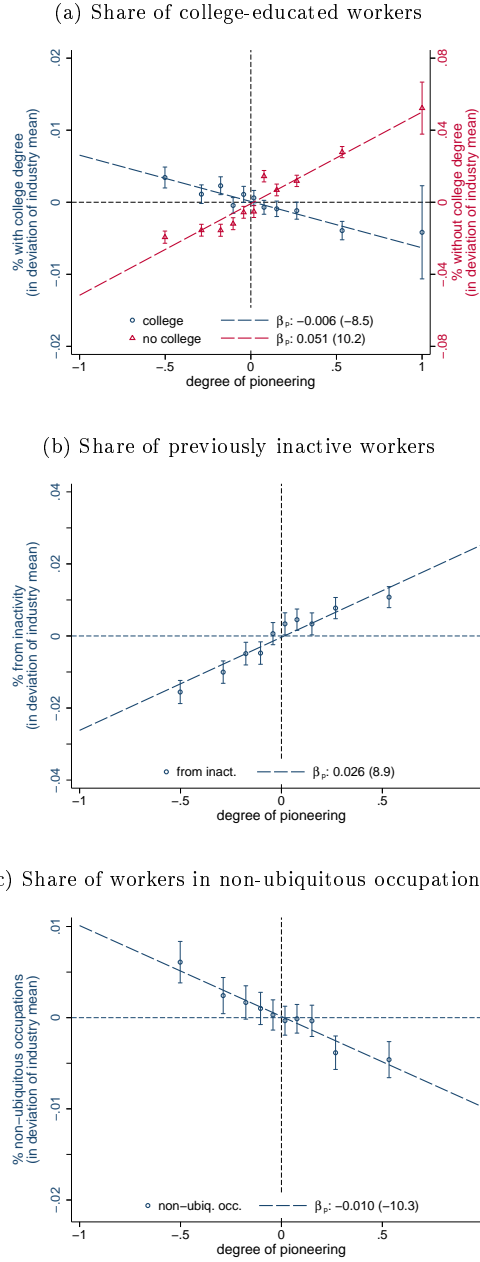
Identifying crucial workers

Pioneers' willingness to hire a disproportionate share of their workforce from outside the region shows how much experienced workers are valued when setting up new plants. However, pioneers typically hire a mix of local and non-local workers. Apparently, prior experience is not required for all workers. This raises the question of who among the workers are so crucial to the success of a new venture that employers are willing to make the effort to hire them from outside the region. To investigate this, we estimate the following equation:

$$reg.\,switch_e = \alpha_t + \gamma_i + \delta_r + \beta_p p_{ir\tau-1} + \epsilon_e \tag{7}$$

where $reg.\,switch_e$ is a dummy that evaluates to 1 if worker e came from outside the region. We estimate (7) for 20 equally-sized wage segments. The resulting estimates for β_p are plotted in Figure 6, together with their 95% confidence intervals.
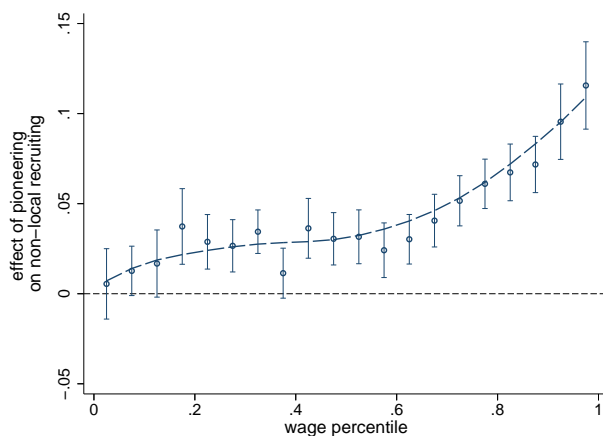
Until about the 60th wage percentile, pioneers barely differ from cluster plants. However, the parameter estimate becomes gradually more negative, increasing to over five-fold in absolute terms as wages approach the 95th percentile. Apparently, although the scarcity of experienced workers in a location forces pioneers to resort to non-local recruiting strategies, they do so only for a relatively small subset of their employees. This means that, in spite of hiring certain indispensable workers from elsewhere, doing so allows pioneers to still

Figure 5: Workforce composition: first-year hires

(a) Share of college-educated workers

(b) Share of previously inactive workers

(c) Share of workers in non-ubiquitous occupations

The figures show the percentage, in deviation of the industry mean, of a certain type of workers among all workers hired by new plants in their starting years for different degrees of pioneering. Vertical lines represent 95% confidence intervals, regression lines are calculated by weighted regression, where weights are given by the inverse of the error variance of the mean. Figure 5a shows the shares of first-year hires with and without a college degree, Figure 5b the share of first-year hires who were previously inactive and Figure 5c the share of of workers in non-ubiquitous occupations.

Figure 6: Effect of pioneering on region switching by wage bin

For this figure all workers hired by new plants are pooled and divided over 20 equally-sized wage bins. Next, a dummy for whether or not a worker entered from a different region is regressed on the new plant's degree of pioneering, controlling for industry, region and year fixed effects. The graph shows the effect of pioneering in each bin, together with a 95% confidence intervals based on standard errors clustered at the industry-region level. The dotted line represents a non-parametric regression smooth of this series.

create a substantial amount of local jobs.

7   The role of western German experts in establishing new industries in eastern Germany

We have analyzed German pioneer plants, starting about 10 years after the reunification of East and West Germany. Historically, East Germany had been the center of many of Germany's manufacturing industries (Braun et al. 2014).[24] Before the Second World War, manufacturing output per capita was 20% higher in what would later become the German Democratic Republic (GDR) than in the territories of the Federal Republic of Germany (FRG). However, with many factories destroyed in the war and firms relocating to West Germany (Buenstorf and Guenther 2007) or dismantled and shipped as war reparations to Russia (Braun et al. 2014), East German manufacturing prowess had been dealt a significant blow. Still, at the fall of the Berlin Wall, East Germany employed about 3 million manufacturing workers (Braun et al. 2014). However, this was an unsustainable level. After Germany's *Treuhandanstallt* broke up 8,500 of East Germany's *Kombinate*, about 15,000 out of the originally tens of thousands economic establishments remained (Braun et al. 2014) and manufacturing collapsed

---

[24]Much of what follows is based on the description of the eastern German economy in Braun et al. (2014).

under the exposure to West German competition, with employment falling by 66% between 1991 and 1997. Only after 1997, eastern German manufacturing employment started recovering, reaching about 50% of its 1991 levels by the year 2008 (DeStatis 20 Jahre Deutsche Einheit).
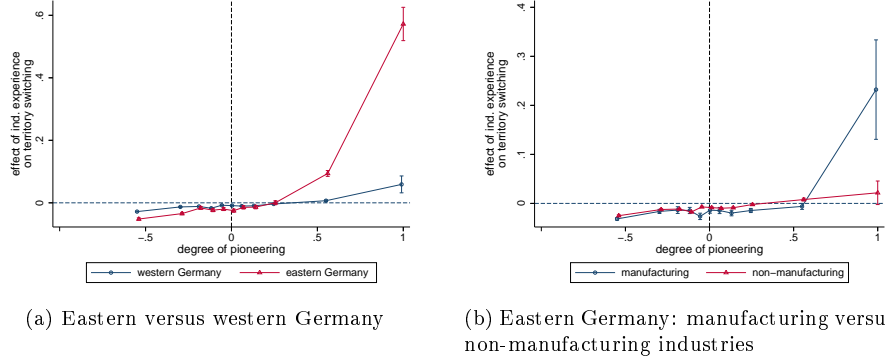
The post-reunification years had thus been marked by dramatic structural change in the eastern German economy. This process was supported through massive subsidies that aimed at modernizing the eastern German capital stock at a rate of up to 50% of total investments in the east (Schmidt 1996). Moreover, because of the collapse of its tradable industries, eastern Germany ran a trade deficit reaching 1,250 Billion DM between 1990 and 1995 largely funded by transfers from the west (Schmidt 1996). Between 1991 and 1998, subsidies, tax breaks and welfare transfers amounted to almost 1.4 Trillion DM (Gerling 2000). However, although these transfers – and the concomitant push by labor unions for rapid wage convergemce to western levels – undoubtedly reduced out-migration, they did not prevent that large numbers of people moved from eastern to western Germany (Hunt 2006). Indeed, by 2006, east-west migration had reached 2.45 million individuals (Fuchs-Schündeln and Schündeln 2009).

Most of the literature – and public debate – on the German reunification have focused on these monetary transfers and on how the east-west migration impacted the economy in both eastern and western Germany (Schmidt 1996, Gerling 2000, Hunt 2006, Uhlig 2006, Burda 2006). In the same period however, 1.45 million individuals had moved in the opposite direction, from the west to the east (Fuchs-Schündeln and Schündeln 2009). In light of our findings so far, this west-east migration may have played a role in transforming the eastern German economy. In particular, the fall of the Berlin Wall did not just provide job opportunities to East Germans and an increased supply of labor to West German firms, but also allowed East German regions access to experienced West German workers they needed to diversify their economy. To understand the importance of this second phenomenon, we study how workers moved between the territories of the former GDR ("eastern Germany") and the former FRG ("western Germany"). In particular, we explore how often experienced workers in pioneer locations are recruited not just from outside the region, but from outside these territories.

Figures 8a and 8b give insight into how important western Germany is as a source of experienced labor for eastern German plants. Figure 8a depicts the difference between the shares of experienced and inexperienced workers who were hired from western territories in the east and vice versa at different degrees of pioneering. The figure shows that clusters in the east hire 5 pp more of their experienced than of their inexperienced workers from within the east (Figure 8a), which is similar to the west, where inexperienced workers are hired from within the west at a 3 pp higher rate than experienced workers are. This starkly contrasts with eastern pioneer locations, where experienced workers come 57 pp more often than inexperienced workers from western Germany. The equivalent number for western pioneer locations (6 pp) is much lower.

Most of the eastern pioneers' demand for western German experienced workers is driven by manufacturing pioneers (Figure 8b). In these industries, experi-

Figure 7: Differences in territory switching between experienced and inexperienced hires

(a) Eastern versus western Germany

(b) Eastern Germany: manufacturing versus non-manufacturing industries

The figures show parameter estimates for different degree-of-pioneering segments of the data in a regression of a territory switching dummy on an experienced worker dummy (defined as a worker who was hired from within the new plant's industry), controlling for year fixed effects. Degree-of-pioneering values represent the median value of a data segment. Figure 8a reports separate estimates for eastern and western Germany. Figure 8b reports results for manufacturing and non-manufacturing industries separately in eastern Germany.

enced workers are hired at a 63 pp higher rate from western Germany than inexperienced workers are. Given that eastern manufacturing pioneers hire about 15% of their inexperienced workers from the west, these pioneers recruit almost four out of five experienced workers from western Germany. By contrast, western German pioneers recruit just one out of four experienced workers from eastern Germany.

To evaluate how much of eastern German pioneers' reliance on western German experienced workers can be ascribed solely to the fact that they are pioneers, holding constant industry, region and other effects, we regress the territory switching dummy on a three-way interaction between variables that distinguish among (1) the new plant's location (eastern or western Germany), (2) whether or not workers are experienced (i.e., hired from within the industry) and (3) the extent to which the new plant is a pioneer. Letting EXP denote workers with industry experience, and E plants in eastern Germany, we arrive at the following regression equation:

$$d_e^{EW} = \alpha_t + \gamma_{i \times E} + \delta_r + \beta_p p_{ir\tau-1} + \beta_{EXP} EXP_e + \beta_E E_e + \qquad (8)$$
$$\beta_{p \times EXP} p_{ir\tau-1} EXP_e + \beta_{p \times E} p_{ir\tau-1} E +$$
$$\beta_{p \times E \times EXP} p_{ir\tau-1} E_e EXP_e + \epsilon_e$$

where d_e^{EW} is a dummy for territory switching. $\gamma_{i \times E}$ represents industry-territory fixed effects. These effects ensure that industries that are found in only one location in the territory do not contribute to the estimates. The three-way-interaction term, $\beta_{p \times E \times EXP}$, quantifies how much more eastern pioneers

rely on western Germany for finding *experienced* workers than western pioneers on eastern Germany, controlling for the overall extent to which eastern pioneers hire western workers.

Table 10 shows the results for four different segments of the data. The lower panel decomposes the three-way interactions into the change in territory switching a one-unit increase in degree of pioneering implies for different subgroups of workers. The three-way interaction term in column (1) shows that, overall, pioneers hire 8.8 pp more of their experienced workers from the west. However, this effect is substantially higher for college-educated workers (15.3 pp) and for manufacturing industries (27.3 pp), reaching 37.7 pp for the intersection of both categories (i.e., college-educated workers in manufacturing industries). This shows that the emergence of new manufacturing industries in eastern German regions was accompanied by a substantial influx of experienced workers from the west.

However, these results could simply reflect that the industries of eastern German pioneers tended to be concentrated in western Germany. In that case, there simply are very few experienced potential recruits in eastern Germany. To control for this, we create a variable that captures which share of an industry's employment was located in western Germany in the year before the new plant entered. We then generate the same three-way interactions with this share as we did for the degree of pioneering. Adding these interactions to equation (8) reduces point estimates of $\beta_{p \times E \times EXP}$ by only between 2 pp (Column 1) and 8 pp (Column 4), representing reductions of at most 21%,[25] showing that our findings are not mainly driven by a relative scarcity of experienced eastern German workers.

Another caveat is that some new establishments in the east may belong to companies that are head-quartered in the west. In principle, this problem is mitigated by filtering out spurious plant openings and spinoffs from existing companies based on the approach by Hethey and Schmieder (2010). However, we cannot exclude that some of the new establishments in our data set are indeed new branch plants of existing western German firms. Indeed, based on the IAB establishment panel, Günther and Gebhardt (2005), report that, in 2001, 15.0% of eastern German manufacturing establishments, together representing 37.8% of manufacturing employment, were owned by western German firms. This highlights the role western German firms played in the renewal of eastern German manufacturing. However, it does not change the fact that many of the plants – be it startups or subsidiaries of existing firms – that pioneered new local manufacturing industries in eastern Germany relied on western Germany for most of their experienced workers. Apart from the massive transfers to help modernize the eastern German capital stock, western Germany therefore also contributed to the economic transformation of eastern Germany because, after the fall of the Berlin Wall, the experience that workers had gained in western industries could freely flow into eastern Germany.

---

[25] Results are reported in the Appendix.

Table 10: Eastern German pioneers' reliance on western German experienced workers

dep. var.:	(1) ind.: all workers: all	(2) ind.: all workers: coll.	(3) ind.: manuf. workers: all	(4) ind.: manuf. workers: coll.
p	-0.002	0.000	-0.001	-0.001
	(0.001)	(0.007)	(0.002)	(0.011)
p x EXP	0.018***	0.026**	0.002	0.014
	(0.003)	(0.010)	(0.005)	(0.017)
p x E	-0.025***	-0.043**	-0.032**	-0.085**
	(0.006)	(0.019)	(0.016)	(0.042)
p x E x EXP	0.088***	0.153**	0.273**	0.377***
	(0.028)	(0.074)	(0.111)	(0.128)
E	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
EXP	-0.007***	-0.006*	-0.012***	-0.009
	(0.001)	(0.003)	(0.002)	(0.010)
E x EXP	0.010*	0.023	0.021	0.038
	(0.005)	(0.019)	(0.019)	(0.044)
constant	0.066***	0.084***	0.070***	0.080***
	(0.002)	(0.008)	(0.007)	(0.027)
year dummies?	yes	yes	yes	yes
ind.-territory dummies?	yes	yes	yes	yes
region dummies?	yes	yes	yes	yes
R-squared	0.071	0.252	0.189	0.472
# obs.	2,570,638	120,126	217,357	17,522
Effect of p→p+1				
east, inexperienced	-0.027	-0.042	-0.033	-0.086
east, experienced	0.079	0.137	0.242	0.305
west, inexperienced	-0.002	0.000	-0.001	-0.001
west, experienced	0.016	0.027	0.001	0.013
Differenced experienced / inexperienced				
west	0.018	0.026	0.002	0.014
east	0.106	0.179	0.275	0.391
Differenced east / west	0.088	0.153	0.273	0.377

The dependent variable is a dummy for whether a worker was hired from an employer located in the other former territory. p-values: ***: .01, **: .05, *:.10, clustered standard errors at the industry-territory level in parentheses.

8 Conclusions

Controlling for region- and industry-specific factors, a low local level of pre-entry specialization in an industry makes it more likely that new plants hire their workers from outside the region. This holds even more so if these workers are college-educated and if the plant operates in a manufacturing industry. The workers that such plants, which are industry pioneers in their region, tend to recruit from other regions typically have industry experience. However, these pioneers also leverage the low-cost advantage of their location by biasing the composition of their workforce towards lower skilled workers for whom industry experience is less relevant.

The willingness of pioneers to attract their best-paid workers from other regions suggests that being able to access experienced workers is an important element in allowing regions to develop new activities. Indeed, within the context of eastern Germany, we find that, even ten years after reunification, western Germany played a key role in providing experienced workers to plants that set up new local industries in the east. This pattern of pioneers hiring experienced workers from the west is most pronounced in manufacturing pioneers, and particularly when hiring college-educated workers.

These findings suggest that human capital is to some extent industry specific and that this industry-specificity slows down the rate at which economic activities diffuse to other locations. This paper has, therefore, implications for the puzzle behind the slow pace of technological diffusion across countries. While Germany has high educational achievement and other positive characteristics of its educational and training systems, the diffusion of industries in its territory involves the movement of workers with industry experience from other regions. If an international border were to restrict such movements, as they did in East Germany after 1948 and even more after 1961, technological diffusion would slow down. The conjecture is that restrictions to international migration would have large effects on the speed of technological diffusion. This is an area that merits further research.

References

Agrawal, A., Cockburn, I. and McHale, J.: 2006, Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships, *Journal of Economic Geography* **6**(5), 571–591.

Almeida, P. and Kogut, B.: 1999, Localization of knowledge and the mobility of engineers in regional networks, *Management science* **45**(7), 905–917.

Bahar, D., Rapoport, H. et al.: 2014, Migration, knowledge diffusion and the comparative advantage of nations, *Technical report*, Harvard University Working Paper.

Barro, R. J. and Sala-i Martin, X.: 1992, Convergence, *Journal of political Economy* pp. 223–251.

Bender, S., Haas, A. and Klose, C.: 2000, Iab employment subsample 1975-1995: opportunities for analysis provided by the anonymised subsample.

Björkman, I. and Gertsen, M.: 1993, Selecting and training scandinavian expatriates: Determinants of corporate practice, *Scandinavian Journal of Management* **9**(2), 145–164.

Bonache, J. and Brewster, C.: 2001, Knowledge transfer and the management of expatriation, *Thunderbird International Business Review* **43**(1), 145–168.

Braun, G., Güra, T., Henn, S., Lang, T., Schürmann, C., Voss, K. and Warszycki, P.: 2014, Atlas der industrialisierung der neuen bundesländer, *Technical report*, BMWi, Berlin.

Breschi, S. and Lissoni, F.: 2009, Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, *Journal of Economic Geography* p. lbp008.

Buenstorf, G. and Guenther, C.: 2007, No place like home? location choice and firm survival after forced relocation in the german machine tool industry, *Location Choice and Firm Survival After Forced Relocation in the German Machine Tool Industry (August 31, 2007). Jena Economic Research Paper* (2007-053).

Buenstorf, G. and Klepper, S.: 2010, Why does entry cluster geographically? evidence from the us tire industry, *Journal of Urban Economics* **68**(2), 103–114.

Burda, M. C.: 2006, Factor reallocation in eastern germany after reunification, *The American economic review* pp. 368–374.

Collins, H. M. and Pinch, T.: 1993, *The golem: What you should know about science*, Cambridge University Press.

Collins, H. and Pinch, T.: 1998, *The golem at large: What you should know about technology*, Cambridge University Press.

Dahl, M. S. and Sorenson, O.: 2009, The embedded entrepreneur, *European Management Review* **6**(3), 172–181.

Dahl, M. S. and Sorenson, O.: 2012, Home sweet home: Entrepreneurs' location choices and the performance of their ventures, *Management Science* **58**(6), 1059–1071.

DeStatis: 20 Jahre Deutsche Einheit, 2010, *Technical report*, Statistisches Bundesamt, Wiesbaden.

DieWelt: 2012, Wie aus einer bier-revolte das erste pils entstand.
URL: *http://www.welt.de/109639795*

Duranton, G. and Puga, D.: 2004, Micro-foundations of urban agglomeration economies, *Handbook of regional and urban economics* **4**, 2063–2117.

Fosfuri, A., Motta, M. and Rønde, T.: 2001, Foreign direct investment and spillovers through workers' mobility, *Journal of international economics* **53**(1), 205–222.

Fuchs-Schündeln, N. and Schündeln, M.: 2009, Who stays, who goes, who returns?, *Economics of Transition* **17**(4), 703–738.

Gartner, H. et al.: 2005, The imputation of wages above the contribution limit with the german iab employment sample, *FDZ Methodenreport* **2**, 2005.

Gerling, K.: 2000, Subsidization and structural change in eastern german transition: Did economic policy meet its objectives?, *Technical report*, Kieler Arbeitspapiere.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. and Shleifer, A.: 1992, Growth in cities, *Journal of Political Economy* **100**(6), 1126–52.

Greenstone, M., Hornbeck, R. and Moretti, E.: 2010, Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings, *Journal of Political Economy* **118**(3), 536–598.

Günther, J. and Gebhardt, O.: 2005, Eastern germany in the process of catching up: The role of foreign and west german investors in technological renewal, *Eastern European Economics* **43**(3), 78–102.

Hausmann, R. and Klinger, B.: 2006, Structural transformation and patterns of comparative advantage in the product space.

Hébert, L., Very, P. and Beamish, P. W.: 2005, Expatriation as a bridge over troubled water: A knowledge-based perspective applied to cross-border acquisitions, *Organization Studies* **26**(10), 1455–1476.

Helsley, R. W. and Strange, W. C.: 1990, Matching and agglomeration economies in a system of cities, *Regional Science and urban economics* **20**(2), 189–212.

Henderson, V., Kuncoro, A. and Turner, M.: 1995, Industrial development in cities, *Journal of Political Economy* **103**(5), 1067–1090.

Hethey, T. and Schmieder, J. F.: 2010, Using worker flows in the analysis of establishment turnover–evidence from german administrative data, *FDZ Methodenreport* **6**(en), 43.

Hunt, J.: 2006, The economics of german reunification, *Unabridged version of entry prepared for the New Palgrave Dictionary of Economics* .

Keller, W.: 2004, International technology diffusion, *Journal of Economic Literature* **42**, 752–782.

MacKenzie, D. and Spinardi, G.: 1995, Tacit knowledge, weapons design, and the uninvention of nuclear weapons, *American journal of sociology* pp. 44–99.

Marshall, A.: 1890, *Principles of economics*, Macmillan London.

McCann, P. and Simonen, J.: 2005, Innovation, knowledge spillovers and local labour markets, *Papers in Regional Science* **84**(3), 465–485.

Mendenhall, M. E., Dunbar, E. and Oddou, G. R.: 1987, Expatriate selection, training and career-pathing: a review and critique, *Human Resource Management* **26**(3), 331–345.

Miller, E. L.: 1973, The international selection decision: A study of some dimensions of managerial behavior in the selection decision process, *Academy of Management Journal* **16**(2), 239–252.

Neffke, F. and Henning, M.: 2013, Skill relatedness and firm diversification, *Strategic Management Journal* **34**(3), 297–316.

Peak, M. H.: 1997, Darned expensive to take for granted, *Management Review* **86**(1), 9.

Polanyi, M.: 1967, The tacit dimension.

Rosenthal, S. S. and Strange, W. C.: 2004, Evidence on the nature and sources of agglomeration economies, *Handbook of regional and urban economics* **4**, 2119–2171.

Sala-i Martin, X. X.: 1996, Regional cohesion: evidence and theories of regional growth and convergence, *European Economic Review* **40**(6), 1325–1352.

Saxenian, A.: 2007, *The new argonauts: Regional advantage in a global economy*, Harvard University Press.

Schmidt, K.-D.: 1996, German unification: A progress report, *Technical report*, Kiel Working Papers.

Uhlig, H.: 2006, Regional labor markets, network externalities and migration: The case of german reunification, *The American Economic Review* **96**(2), 383.

Appendix

Table A1: Triple interaction with western German industry shares

dep. var.:	(1) ind.: all workers: all	(2) ind.: all workers: coll.	(3) ind.: manuf. workers: all	(4) ind.: manuf. workers: coll.
p	-0.002	0.001	-0.001	0.003
	(0.001)	(0.007)	(0.002)	(0.011)
p x EXP	0.016***	0.024**	0.000	0.000
	(0.003)	(0.011)	(0.004)	(0.011)
p x E	-0.022***	-0.038*	-0.026**	-0.077*
	(0.006)	(0.020)	(0.013)	(0.041)
p x E x EXP	0.070**	0.124	0.235***	0.298***
	(0.028)	(0.082)	(0.087)	(0.111)
sh(WG)	-0.154***	-0.361***	-0.101	0.088
	(0.030)	(0.114)	(0.109)	(0.342)
sh(WG) x EXP	-0.026**	-0.025	-0.053	-0.452
	(0.011)	(0.026)	(0.058)	(0.361)
sh(WG) x E	0.021	0.376	0.224	3.509
	(0.068)	(0.296)	(0.362)	(2.243)
sh(WG) x E x EXP	0.143***	0.179	0.478*	1.447**
	(0.044)	(0.131)	(0.271)	(0.718)
E	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
EXP	0.014	0.014	0.033	0.394
	(0.009)	(0.020)	(0.051)	(0.327)
E x EXP	-0.103***	-0.118	-0.379*	-1.214**
	(0.032)	(0.098)	(0.215)	(0.615)
constant	0.185***	0.301***	0.114	-0.649
	(0.021)	(0.088)	(0.098)	(0.474)
year dummies?	yes	yes	yes	yes
ind.-territory dummies?	yes	yes	yes	yes
region dummies?	yes	yes	yes	yes
R-squared	0.071	0.252	0.189	0.475
# obs.	2,570,638	120,126	217,357	17,522

The dependent variable is a dummy for whether a worker was hired from an employer located in the other former territory. p-values: ***: .01, **: .05, *:.10, clustered standard errors at the industry-territory level in parentheses. sh(WG) represents the share of western Germany in an industry's previous year's total German employment.