Paper to be presented at the

35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19

# The workforce of pioneer plants

**Ricardo Hausmann**
Harvard University
Kennedy School of Government
ricardo_hausmann@harvard.edu

**Frank Neffke**
Harvard University

frank_neffke@hks.harvard.edu

**Anne Otto**
Institut für Arbeitsmarkt- und Berufsforschung

Anne.Otto@iab.de

Abstract

How do plants get their workforce when they are industry pioneers in a particular location? By definition, these plants can either hire local workers without prior industry experience or experienced workers from outside the region. We investigate to what extent pioneer plants trade off geographic distance against industry experience when recruiting their workers and how they differ in this from their peers inside the industry?s main agglomerations. Using a dataset on all individuals that contribute social security payments in Germany, we find that plants outside industry agglomerations recruit workers both from more distant regions and from less related industries. The exact balance between geographic and skill distance depends on the type of worker and on the type of industry in which the plant operates. For highly skilled workers and for workers in industries that produce tradable products, the balance tilts towards pioneer plants recruiting from farther away, not from less related industries, than plants inside clusters. The opposite holds for low skilled workers and workers in plants producing non-tradable products.

Jelcodes:R11,O12

# The workforce of pioneer plants

Preliminary version – please do not quote

**May 2013**


## Abstract

How do plants get their workforce when they are industry pioneers in a particular location? By definition, these plants can either hire local workers without prior industry experience or experienced workers from outside the region. We investigate to what extent pioneer plants trade off geographic distance against industry experience when recruiting their workers and how they differ in this from their peers inside the industry's main agglomerations. Using a dataset on all individuals that contribute social security payments in Germany, we find that plants outside industry agglomerations recruit workers both from more distant regions and from less related industries. The exact balance between geographic and skill distance depends on the type of worker and on the type of industry in which the plant operates. For highly skilled workers and for workers in industries that produce tradable products, the balance tilts towards pioneer plants recruiting from farther away, not from less related industries, than plants inside clusters. The opposite holds for low skilled workers and workers in plants producing non-tradable products.



**Ricardo Hausmann**

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

**Frank Neffke**

Harvard Kennedy School, Center for International Development, Harvard University

**Anne Otto**

Institut für Arbeitsmarkt- und Berufsforschung

# 1. Introduction

How does regional diversification deal with the obvious fact that pioneers into a new industry cannot hire local workers with industry experience? How important is human capital specificity in constraining entry into new industries at the regional level? We investigate these questions by comparing how plants in industries that are new to a region source their workers in comparison with new plants in regions that already have an ample presence of other firms in the same industry. In principle, if a region hosts many firms of a given industry, it is reasonable to assume that it would tend to have thick markets for the specialized inputs the industry requires. However, if an industry is new to a location, plants will not always be able to find the labor and other inputs that match their requirements. Whereas many intermediate inputs are tradable and can thus be imported from elsewhere, labor is much less mobile. Because the jobs often require tacit knowledge that can only be acquired through work experience in an industry (Polanyi, 1967), the absence of other plants in the industry constitutes a challenge. Indeed, 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 may form an important impediment to economic development. In the words of Marshall (1890, IV.X.9): "[t]hese difficulties[1] 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." In absence of other plants in the industry, a pioneer plant has only two choices. Either it hires workers without industry experience or it hires workers that gained this experience elsewhere. In this paper, we study how the absence of other plants in the own industry affects the ways in which new plants assemble their workforces.

The role of the local labor market in economic development has received substantial attention in the literature on agglomeration externalities (Marshall 1890, Glaeser et al. 1992, Henderson et al. 1995), local clusters (Porter 1998, Porter 2003) and regional innovation systems (Cooke and Morgan 1998). Moreover, the available pool of labor has been put forward as an important factor in explaining why regions diversify into activities that are closely related to existing ones (Frenken and Boschma, 2009). We aim to shed light on these issues by studying where the industry pioneers of a region recruit their

[1] 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, in the remainder of this quote, Marshall also expresses his expectation that these problems were "however being diminished by the railway, the printing-press and the telegraph," something which this paper will put in doubt.

labor. To what extent do plants resort to workers from outside their regions and to what extent can experienced workers be substituted by others? Moreover, we argue that the way new plants deal with the absence of experienced workers in their region will differ by industry and worker types. To motivate this, we set up a simple model in which plants face uncertainty about the productivity of inexperienced workers. In this model, the only way for pioneer plants to overcome this uncertainty is by hiring experienced workers from outside the region. Whether it is optimal for a plant to hire local or non-local workers depends on how strongly productivity is hit when a low quality worker is hired. The more critical the job of a worker is to the functioning of the plant, the larger the productivity loss will be if the firm hires a low productivity worker. However, the recruitment strategy of a plant will also depend on whether it can easily absorb productivity losses. 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. Therefore, the less tradable the products of a pioneer plant, the more it can afford the risk of hiring low quality workers. As a consequence, our model predicts that both crucial workers and workers in tradable industries are more likely to come from related industries, but from other regions.

We put these ideas to the test, using individual-level data from the German social security records that cover about 80% of the German labor market and study the workforce of new plants in Germany between 1998 and 2008. Although almost all new plants are found in regions with large agglomerations of the plants' industries, some plants locate outside such clusters. As expected, these pioneer plants typically differ from plants inside clusters in the origins of their workers. Workers in pioneer plants more often come from farther away and from less related industries. Moreover, before entering the new plant employees more often worked in a different occupation or were not active on the labor market than workers in plants in cluster locations. Moreover, in tradable industries, pioneer plants differ from cluster plants more in terms of the geographic distance over which they recruit their workers than in terms of how related the industries are from which these workers come. The opposite holds for plants that operate in non-tradable industries: compared to their peers inside clusters, pioneer plants in non-tradable industries tend to hire workers from less related industries rather than from outside their regions. Similar differences exist among worker types. Pioneer plants hire particularly often key workers (i.e., highly paid, highly educated or with rare skills) from farther away instead of from less related industries, whereas the reverse holds for less crucial, lower skilled workers. This suggests that regional pioneers can hire some workers with less well-matched skills, but key personnel needs to possess the right experience.

The fact that plants are observed to hire key personnel from outside the region whenever experienced workers are locally unavailable, suggests that the skills in the local labor market limit a region's scope for diversification and, given the low mobility of workers across country borders, explains why a fortiori the same should hold for countries (Hausmann and Klinger, 2007). It also explains why multinationals are important in diffusing skills across the globe (Fosfuri et al. 2001; Keller 2004) and why they have to 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 (Breschi & Lissoni 2009, Agrawal et al. 2006). Whereas these authors show how inventor mobility diffuses knowledge across regions, we show that labor mobility is required to allow industries to diffuse to other regions. Finally, our work is also related to the literature on the home bias of entrepreneurs (Dahl and Sorenson 2009, 2012, Klepper and Buenstorf 2009), but, with the exception of Timmermans (2010), this literature focuses 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 in from where they recruit their workers.

In section 2, we provide an overview of the related literature. Section 3 derives a model for the geographical labor recruitment strategy of pioneer plants. Section 4 discusses our data source and research strategy. Section 5 presents the empirical findings and section 6 concludes.

## 2. Tacit knowledge and regional diversification

Economic convergence is a slow process at best. This holds as much for the convergence of countries as for regions within countries. Slow convergence is commonly ascribed to the fact that, for local economies to converge in income levels, the superior technologies employed in advanced economies need to diffuse to lagging economies. Moreover, recent research shows that a country's opportunities for diversification are strongly dependent on the current mix of activities in an economy. Evidence for such path dependent diversification is provided by Hausmann and Klinger (2006) and Hidalgo et al. (2007), who find that nations diversify their exports predominantly into products that are closely related to their current export basket. The same holds for the diversification of the industrial portfolios of Swedish regions (Neffke et al., 2011) and of export baskets of Spanish regions (Boschma et al., 2013). Why however, in an age of almost costless and instantaneous communication, do industries and their technologies not diffuse more rapidly?

A common response to this question in economic geography and international economics is that all knowledge has a tacit component (Polanyi, 1958). That is, for many of our capabilities, we are unable to verbally explain how they work. As a consequence, we can communicate only a small fraction of what we know and learning a skill often involves observing and collaborating with people who have mastered it. 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, apart from a deep understanding of nuclear physics, nuclear weapons design also involves much apprenticeship learning and expert judgment (MacKenzie and Spinardi, 1995). Access to tacit knowledge is therefore considered a *sine qua non* for firms to survive in almost any industry.

Due to its tacit nature, much knowledge is embedded in the experience and skills of workers. Therefore, new plants cannot simply buy all the technology they need but have to acquire knowledge by hiring workers with the right skills. That is one reason why foreign multi-national enterprises (MNEs) reassign experienced employees to their new branch plants abroad (Bonache and Brewster 2001). Given the costs of such expatriation, the very practice shows that companies understand the importance of employee mobility to transfer knowledge. Interestingly, Hébert et al. (2005) report findings that suggest that expatriation is particularly successful when industry specific knowledge needs to be transferred. Indeed, the selection of expatriates is based predominantly on candidates' technical, not cross-cultural skills (Miller 1973; Mendenhall et al. 1987; Björkman and Gertsen 1993; Bonache and Brewster 2001).

However, not all plants have parent firms that provide them with an internal labor pool from which to hire specialized workers. Instead, these plants have to rely mostly on local labor markets. Therefore, a specialized local labor force may generate strong externalities (Marshall 1890, Glaeser 1992, Henderson et al. 1995). On the one hand, these so-called localization externalities are generated because thicker labor markets create higher quality matches between employers and employees (Helsey and Strange 1999; Duranton and Puga 2004). On the other hand, the labor market and labor mobility are an important channel for 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 knowledge that will be created in the future.

Local labor markets are also important because they spawn the entrepreneurs needed to set up new plants. Often, these entrepreneurs set up firms in their home regions. This home bias may be due to personal reasons (Dahl and Sorenson 2009), but also due to more tangible benefits. For instance, Dahl and Sorenson (2012) find that “region tenure,” i.e., the number of years someone has worked in the region, increases the survival chances by almost as much as industry tenure does. Dahl and Sorenson attribute this effect partly to entrepreneurs’ know-who: local entrepreneurs can draw on their local networks to find employees. This again suggests that local labor markets are important for firms in gaining access to skills and know-how.

In sum, for regions to diversify, some plants must enter new industries. If finding the right workers is indeed an important hurdle that new plants need to overcome, then the success of a firm depends in part on the local availability of these workers. In particular plants in industries that are new to the region will struggle because of the lack of thick, specialized local labor markets. As a consequence, locating far away from the hotspots of an industry may severely limit the access to this labor. An important explanation for the lack of diffusion of technologies and industries may therefore be the fact that labor is not perfectly mobile across regions. To shed light on this issue, we investigate the origins of the workforce of pioneer plants. To the extent that pioneer plants depend more than other plants on workers that are willing to relocate from other regions, geographic labor mobility will play a role in the ability of regions to diversify.

## 3. A model of pioneer plant’s labor recruitment

When deciding whom to hire, a plant needs to balance the costs and benefits of hiring a particular worker. Workers with prior experience that fits the needs of a plant will, presumably, be more productive. The more related the industry in which someone previously worked, the more productive she will be in the new plant. Because new plants that are industry pioneers in a particular region will not have a local labor market with workers that have much industry experience, they will face the choice of hiring local workers with less industry experience or looking for workers in other regions. This requires convincing workers to relocate, for which they often have to be compensated, not just for the outright costs of relocation but also for the fact that the shallowness of the local labor market will increase their hold-up risks. Relocation costs may rise with the distance between the old and the new job. A pioneer plant thus has to balance the benefits from hiring workers with the right skills with the costs that it incurs to recruit them from far away.

In which direction this balance will tilt, will depend on the type of worker, but also on the type of industry the plant belongs to. Some workers are more crucial to the success of a plant than others. For instance, hiring a competent assembly line worker may be more critical than hiring a cleaner. Moreover, some jobs demand highly industry specific skills, whereas other jobs may require more general skills. For instance, accounting clerks may have less industry specific human capital than engineers. The more critical and specific the skills a job requires, the more important it is to hire workers with the right prior experience.

Apart from the type of worker, the type of industry in which a plant operates may play a role as well. The geographical extent of a plant's market depends, among other things, on its industry. Given that they are the first to enter a region, pioneer plants have some monopoly power in non-tradable industries. In tradable industries in contrast, pioneer plants face competition from plants in other regions. Therefore, they can less afford experimenting with workers whose experience does not clearly signal that they have the skills required for their new jobs. Moreover, tradable industries may also more often produce complex products that require highly specific skills. Both due to higher complexity products and due to stronger competitive pressures, pioneer plants in tradable industries will tend to opt for workers with closely related industry experience, even if that means recruiting them from far away.

We synthesize the above considerations in a model in which pioneer plants choose between hiring workers from outside ("foreign" workers) or inside their regions ("domestic" workers). By definition, pioneer plants face no competition from within their region. Therefore, the less tradable the products of pioneer plants, the more pricing power they have. At the same time, however, pioneer plants will be unable to find experienced workers in their regions. As a consequence, they must either hire inexperienced domestic workers, or experienced foreign workers. The problem of hiring inexperienced workers is that it is uncertain whether they dispose of the right skills for the job. This makes hiring inexperienced workers risky.

In our model, plants are small and produce one unit of output. Their only input is labor. Labor can either be adequate (H-type workers) or not (L-type workers). If a worker previously worked in the pioneer plant's industry, her quality will be H. However, if she worked in another industry, her quality is unknown. In the latter case, with probability $\lambda$ she will be an H-type worker, with probability $(1-\lambda)$ an L-type worker. A plant can avoid the risk of ending up with an L-type worker by hiring a foreign worker with the right industry experience, but then it has to pay a relocation fee equal to $R$. We assume that relocation costs do not depend on a worker's human capital.[2] Wages are fixed by the alternative employment opportunities of workers in the rest of the economy. These wages are the same for all workers. Finally, if a plant runs negative profits, it faces additional bankruptcy costs that are proportional to the incurred losses, where the coefficient of proportionality is $\theta$. Given all this, a plant's expected costs are:

With foreign workers:

$$C = w + R \tag{1}$$

With domestic workers:

$$C = \begin{cases} \lambda w + (1-\lambda)\frac{w}{q} & if\ \frac{w}{q} - P \leq 0 \\ \lambda w + (1-\lambda)\left[\frac{w}{q} + \theta\left(\frac{w}{q} - P\right)\right] & if\ \frac{w}{q} - P > 0 \end{cases} \tag{2}$$

In (2), $q$ is a quality parameter of the worker, with $0 < q < 1$. It captures the fact that L-type workers are less productive and hence that a plant requires $\frac{1}{q}$ times the amount of L-workers to do the same work as a mass of one of H-workers. The smaller $q$ is, the greater the difference in quality between L and H workers. $\theta\left(\frac{w}{q} - P\right)$ represents the bankruptcy costs, which are paid only if wages exceed revenues, $\frac{w}{q} > P$.[3]

Plants will prefer domestic over foreign workers if (2) is smaller than (1). This yields the following inequality:

$$\begin{cases} q > 1 - \frac{R}{(1-\lambda)w + R} & if\ \frac{w}{q} - P \leq 0 \\ P > \frac{w}{q} + \frac{1}{\theta}\left[\frac{w}{q} - \left(w + \frac{R}{1-\lambda}\right)\right] & if\ \frac{w}{q} - P > 0 \end{cases} \tag{3}$$

The first and the second inequality in (3) cross at the intersection with the bankruptcy inequality $\frac{w}{q} > P$. Above and to the right of the line that demarcates these inequalities, plants will prefer domestic

[2] The empirical literature shows that high-skill workers tend to be more mobile than low-skill workers (Long 1973), but we take this to be a market outcome, not a characteristic of the worker.

[3] A plant will not produce at all if revenues do not even exceed costs with a good draw, therefore, bankruptcy costs are added only in the scenario of a bad draw.

workers, below that line, they will prefer foreign workers. An increase in $\lambda$ pushes the inequalities in (3) downward and to the left, making domestic workers more attractive at lower values of $P$ and $q$.

However, plants do not produce at all if expected costs exceed revenues. For production with foreign workers, this implies:

$$P > w + R \tag{4}$$

For production with domestic workers, the condition is:

$$P > \begin{cases} \lambda w + (1-\lambda)\frac{w}{q} & if\ \frac{w}{q} - P \leq 0 \\ \frac{\lambda w + (1-\lambda)(1+\theta)w/q}{1+(1-\lambda)\theta} & if\ \frac{w}{q} - P > 0 \end{cases} \tag{5}$$

Again, the equalities move downwards as $\lambda$ increases if $0 < q < 1$.

Figure 1 summarizes the model, indicating in which zones production is feasible and in which domestic workers are preferred over foreign workers.[4]

Figure 1 about here

Domestic workers are preferred over foreign workers when the quality of L-type workers is closer to that of H-type workers (high $q$) and when the plant has higher pricing power (higher $P$). Moreover, an increase in the probability of drawing an H-type worker (higher $\lambda$) increases the attractiveness of domestic production at all levels of $q$ and $P$.

In terms of observables, high levels of $P$ occur when products of a plant are highly non-tradable. Small differences between L- and H-type workers would arise when a worker's job does not require much (industry-specific) human capital. High levels of $\lambda$ can be associated with a situation in which domestic

[4] Note that bankruptcy risk forces plants to suspend their production sooner. Therefore, the following inequality holds:

$$\lambda w + (1-\lambda)\frac{w}{q} < \frac{\lambda w + (1-\lambda)(1+\theta)w/q}{1+(1-\lambda)\theta}$$

Moreover, it is easy to show that the bankruptcy line also lies below the minimum production line for all $0 < q < 1$:

$$\frac{\lambda w + (1-\lambda)(1+\theta)w/q}{1+(1-\lambda)\theta} < \frac{w}{q}$$

workers can be sourced from strongly related industries, which increases the likelihood that they dispose of the right skill set.

In short, the model would predict that, ceteris paribus, jobs in tradable industries (lower P), jobs that require more industry specific knowledge (lower $\lambda$), and jobs for which the productivity between H- and L-type workers is large, will be done by foreign workers.

# 4. Data

## The HES data set

As a data source we use the Historic Employment and Establishment Statistics (HES)[5] database. The HES data are based on Germany's social security records and contain information on daily wages, a range of socio-demographic variables (such as educational attainment, gender, and age) and the industry, occupation, and location where each individual works. The wage information is very reliable, since it is used to determine social security contributions but wages are censored due to the contribution limit to social security. We deflate wages to 2005 prices using the consumer price index from Germany's statistical office. To be represented in the HES, individuals must be subject to Germany's social security system. Consequently, the data set excludes all those who are exempt from social security payments, such as civil servants and self-employed individuals, representing about 20% of the German labor force. Moreover, we exclude all individuals in training and in part-time jobs. All in all, this leaves a data set with about 20 million workers a year. People 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 1998-2008. The original dataset is organized in employment spells. To reduce the computational burden, we convert the data into a longitudinal format with one record per individual per year, corresponding to its employment on June 30.

## Regions and industries

The finest level of geographical detail in our version of the HES is the district level (*Kreis* in German). There are over 400 districts in Germany. Districts are nested within spatial planning regions (so-called *Raumordnungsregionen*), of which there are 96. Henceforth, we use these spatial planning regions as

[5] See Bender, Haas and Klose (2000) for a detailed description of this database.

our definition of regions. However, distances between the current and previous jobs of workers are based on the road distance between the center points of the corresponding districts.

As an industry classification system we use the German *Klassifikation der Wirtschaftszweige* 2003 (WZ 203) at the 5-digit level. At the 4-digit level, this classification matches the European NACE 1.1 classification. In the years 1998-2002, industries are classified according to the *Klassifikation der Wirtschaftszweige* 1993 (WZ 1993), which matches the NACE 1.0 classification at the 4-digit level. We harmonize the 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. All in all, we end up with over 1,000 different 5-digit industries.

## Identifying new plants

The HES establishment identifiers are in some ways 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, for a firm with 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. When we talk about establishments in this paper, these caveats should be taken into consideration.

More problematic is that, if a new establishment identifier appears, this does not unambiguously imply that the establishment is new. Although once assigned however, identifiers should not change if the establishment relocates or is temporarily closed, Hethey and Schmieder estimate that between 35 and 40 percent of all appearing identifiers are in fact not due to the entry of completely new establishments. Establishments may be given new identifiers when they are taken over by other firms or when the firm changes the way it reports to the employment agency. To overcome this problem, we use a similar method as in Hethey and Schmieder (2010) and look at the extent to which an establishment's first workers all come from the same other plant. Our goal is not so much to identify all new plants in the economy as to weed out all dubious new plant entries. Therefore, our selection criteria are more strict than the ones used by Hethey and Schmieder. To qualify as a new plant we require that it fulfills three conditions. First, no more than 25 percent of the employees that are hired within the first three years of a plant's existence may come from the same other plant. Second, in order for this percentage to make sense, we require that the plant hires at least ten employees within these three years. Third, the absolute number of workers that come from the same plant may not exceed 20, regardless of the plant's size. These criteria will ensure that we are really investigating new plants. However, because of the size requirement and the fact that firms must survive for at least three years to enter our selection procedure, we also introduce a selection bias.

## Research strategy

For each of the selected plants, we determine to what extent the plant is a pioneer plant or a cluster plant. We do this by calculating whether the plant's industry is overrepresented in the region in the year prior to the plant's foundation. In line with the literature on international trade (Balassa, 1986), we refer to this as the region's revealed comparative advantage in the industry. In economic geography, the term location quotient is more commonly used. Let $i$ represent the new plant's industry and $r$ the region in which it is located:

$$RCA_{ir} = \frac{emp_{ir}/emp_{.r}}{emp_{i.}/emp} \tag{6}$$

In (6), $emp_{ir}$ is the location in region $r$ in industry $i$ in the plant's pre-entry year. Omitted indices indicate summation over the corresponding class. The RCA is strongly right-skewed. Values between 0 and 1 indicate that the region is underspecialized in the industry, whereas values between 1 and infinity signal overspecialization. Because RCA often takes on a value of zero, we cannot simply diminish the skewedness of the distribution by taking the logarithm of RCA. Instead, we use the following transformation to obtain a more symmetrically distributed variable:

$$RCA_{ir}^{*} = \frac{RCA_{ir}-1}{RCA_{ir}+1} \tag{7}$$

Values corresponding to underspecialization are mapped onto the interval -1 to 0 and values that indicate overspecialization onto the interval 0 to 1. The main variable of interest in this study is the pre-entry $RCA^*$ of the industry in which a new plant is established. It is defined at the level of industry-region cells. Pioneer plants will have a pre-entry $RCA^*$ of -1. However, we will use this term somewhat loosely to indicate plants that locate in industry-region cells with RCA values. The higher the pre-entry $RCA^*$ of a plant's industry-region cell, the stronger the activity is clustered in that location. We will often contrast pioneer plants with those that locate in high pre-entry $RCA^*$ locations. The latter, we will refer to as *cluster plants*.

Next, we look up the origins of all workers that enter a selected plant's workforce between 1999 and 2008. For each of these workers, we track the location, the industry and the occupation in which they worked in the year before they were hired by the new plant. This allows us to define the "distance" between a worker's job in the new plant and his or her previous job in a number of dimensions. First, we use dummy variables that indicate whether a worker previously worked in a different (1) industry ("industry switchers"), (2) planning region ("region switcher") or (3) occupation ("occupation switcher"). We use a fourth dummy for people that were not in our data set in the year before they enter the plant. This may have several reasons. For some of these workers, the new job is their first job, while others may come from abroad, may have been unemployed or marginally employed, on maternity leave, taking additional training etc.. We will refer to these workers as "previously inactive" on the labor market.

We also use two continuous distance variables. The first one is the road distance in kilometers between the two jobs. We define this distance as the distance between the centers of the districts in which the jobs are located. For jobs that are located in the same district, we use the distance to the nearest other district and divide it by two.[6] The second distance measure is supposed to capture the skill distance between two industries. For this we use a variant of Neffke and Henning's (2012) skill relatedness. These authors use labor flows between industries as an indication of similarity in skill requirements in industries. The underlying idea is that workers will predominantly move between industries that require similar skills to avoid making obsolete the human capital they acquired through learning by doing. Let $F_{ij}$ be the total number of workers who moves from industry $i$ to industry $j$. 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:

[6] Note that we use districts, the finest grained regional classification in our data, to calculate distances. The regional unit is however the higher-order planning region.

$$SR_{ij} = \frac{F_{ij}}{F_{i.}F_{.j}} F \quad (8)$$

Values between one and infinity indicate that the labor flow between two industries exceeds what one would expect if the flows were random, considering the overall in- and outflows of workers in the industries. In line with Neffke and Henning, we call these industries skill related. Values between 0 and 1 indicate that industries are unrelated in terms of their skill requirements. Analogously to the transformation of RCA, 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:

$$SR_{ij}^{*} = \frac{SR_{ij}-1}{SR_{ij}+1} \quad (9)$$

In the remainder, we use the average of $SR_{ij}^{*}$ across all years between 1998 and 2008 to measure the relatedness between a worker's old job and her new job. The skill relatedness for workers that come from the same industry is unknown and we will ignore these workers in our analyses of skill relatedness.[7]

The HES also contains information on an individual's daily wage. Wages are right-censored, due to the fact that social security payments are capped at maximum value that changes from year to year. We use the consumer price index provided by Germany's statistical office to deflate all wages, with 2005 as the base year.

We expect that there will be large differences in these variables across industries. Therefore, we normalize all variables using information from all individuals that do not enter any of the new plants. We normalize the dummy variables by subtracting the mean of the variable in the industry. The continuous variables are normalized by subtracting industry means and dividing by industry standard deviations. Means and standard deviations are calculated based on all people who switch jobs from one establishment to another in a particular year, excluding those that enter one of our new plants. The only exception is the mean and standard deviation of wages: here we use all workers in an industry, regardless of whether they switch jobs or not. To increase precision, we drop plants in industries for which mean and standard deviations are based on fewer than 100 observations. For the wage variable, we use information on all individuals in other plants, not just job switchers. Normalizing the RCA*

[7] Alternatively, one could set the transformed relatedness to 1 in these cases.

variable would complicate determining whether a plant is a pioneer plant or not. Therefore, we keep this variable as it is.

## Industry and worker groupings

We group workers based on worker characteristics and plants based on industry characteristics. This allows us to study whether the differences in recruitment strategies between high and low RCA plants differ by worker and industry type. To assess the tradability, we classify industries according to two criteria. On the one hand, we distinguish industries in manufacturing from those in other sectors of the economy, assuming that manufacturing industries tend to be more tradable than other industries. On the other hand, we relate the tradability of an industry to its ubiquity. Non-tradable industries need to locate near final consumers and hence will have to be geographically more dispersed or ubiquitous. We determine the ubiquity of an industry by calculating how far the regional employment shares of an industry deviate from the average national employment share of that industry:

$$U_i = \sum_r \left(\frac{emp_{ir}}{emp_{i.}} - \frac{emp_{.r}}{emp}\right)^2 \tag{10}$$

We average $U_i$ across the years 1998 to 2008 and define the 50% of industries with the lowest $U_i$ values as ubiquitous.

Both the manufacturing/non-manufacturing and the ubiquitous/non-ubiquitous dichotomies aim at distinguishing industries that mainly produce for local markets from those that compete on national or even global markets. In that sense, manufacturing and non-ubiquitous industries are typically “export” industries that sell their goods across regional borders. As argued in section 2, such industries differ from industries that service local markets in the degree of competition they face from other regions and hence in their pricing power. Whereas the latter will often choose a location based on the local demand for their goods and services, the former will have to choose a location that gives them access to the capabilities they require to compete on (inter)national markets.

Workers are grouped according to their wages, according to the highest level of education they acquired, and according to the ubiquity of their occupation. An occupation’s ubiquity is defined analogously to the ubiquity of an industry, only that now we use regions’ occupational shares instead of industrial employment shares in equation (5). Wages are divided into three groups: employees with wages in the lowest quartile of the income distribution of all workers that enter new plants, those with wages in the highest quartile and those in between. The educational classification consists of four classes. At the lowest level, workers have either no high school diploma or a high school diploma without any additional professional education,[8] at the medium level, workers either have a high school diploma with professional education or an Abitur[9] with or without additional professional education. At the highest level, workers have either completed an education at a vocational college (*Fachhochschule*) or at a university (*Hochschule* or *Universität*).

The different worker groupings aim to distinguish between core and non-core workers. The workers that are most crucial to a plant's successful operation are often highly paid and typically also have higher levels of education than other workers. Similarly, workers in non-ubiquitous occupations are more likely to have rare skills that are hard to find in other workers.

## 5. Findings

Tables 1 and 2 provide descriptive information on our dataset. The top panel of Table 1 shows that over the course of 1999 to 2005, almost 20,000 new plants met the selection criteria to be regarded as unambiguously new. By far the largest percentage of these new plants are set up in tradable (i.e., manufacturing or non-ubiquitous) industries. The bottom panel of Table 1 shows how many distinct workers entered these plants between 1999 and 2008. These add up to about 1.7 million workers.[10]

Table 2 gives an overview of the variables we use. The figures in this table are based on all workers in all industries. We report two different rows for pre-entry RCA*. The first row refers to the average RCA* in any industry-region cell in Germany. The numbers in the second row are weighted by the number of workers who enter new plants in each cell, providing an average pre-entry RCA* for new plant activity. About 62% of all workers in new plants did not belong to the labor force as defined in section 3 in the year before they were hired. Of those that did, about 75% switched regions, 43% switched industries and 60% switched occupations.

Table 1 about here

[8] The German school system has different types of high schools. Low education levels here refer to either no high school diploma or a diploma from the "Volks-Hauptschule, Mittlere Reife oder gleichwertige Schulbildung ohne Berufsausbildung," which means that the person in question did not take any vocational education and generally will not be allowed to enter into a university without further schooling that awards them the so-called *Abitur*.

[9] See footnote 8.

[10] Because some workers leave again, this is not the number of employees in new plants in any given year.

Table 2 about here

## New plants' pre-entry RCA

In line with Head et al. (1995), Rosenthal and Strange (2003) and Buenstorf and Klepper (2009), we find that most plants are set up in regions where their industry is already well-represented, locating in industry-region cells that typically have high RCAs. The average RCA* across all industry-region combinations in the German economy is -0.412 (see Table 2). Figure 2 shows the deviation from this average for the region-industry cells in which at least one new plant, distinguishing among plants of different types.[11] Plants enter location-industry cells with RCA* values that are on average about 0.4, or one standard deviation, above the overall average of RCA*. Although plants in manufacturing and non-ubiquitous industries seem to locate in locations with lower RCA* values, the 95% confidence intervals suggest that location choices are fairly similar for plants in different types of industries.

Figure 2 about here

## Post-entry skill fit

Does the workforce of pioneer plants differ from the one in cluster plants? And does this difference decrease over time? We investigate these questions by assessing the extent to which the occupational mix of a new plant's workforce resembles the average occupational mix in the industry. Let $o_p$ be a vector of length $O$, with $O$ the number of different occupations in the economy. Vector $o_p$ contains the occupational employment shares for plant $p$. Similarly, let $o_{I(p)}$ be the vector of occupational employment shares in $p$'s industry $I(p)$. We interpret $o_{I(p)}$ as the ideal occupational mix of the industry and define the *skill fit* as the Pearson correlation between $o_p$ and $o_{I(p)}$:

$$skill\_fit_p = Corr(o_p, o_{I(p)})$$

The higher this number, the more closely plant $p$'s occupational mix resembles the one of its industry. Figure 3 shows how the skill fit evolves with plant age by displaying the average correlation between $o_p$ and $o_{I(p)}$ together with a 95% confidence interval.

Figure 3 about here

[11] We calculate this number as the average pre-entry RCA* value of cells with at least one entry in a year and then average this number across all years.

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 plants' skill fit on their pre-entry RCA* levels at ages 0 to 9, while controlling for industry, region and year fixed effects. Given that large plants allow for more division of labor, we also run regressions that include the logarithm of the plant's entry employment as a control.[12] Plants that enter high RCA locations start out with an occupational mix that resembles the ideal occupational mix more closely than plants in low RCA locations. However, this difference fades out as the plants grow older.[13] This shows that, although pioneer plants start at a disadvantage, they tend to close the gap over time.

Figure 4 about here

## Origins of the workforce

By definition, pioneer plants cannot hire employees from other local firms in their industry. Therefore, pioneer plants can either recruit workers from outside the region or from outside their industry. If we think of pioneer plants as plants in low pre-entry RCA* locations and of cluster plants as plants in high pre-entry RCA* locations, we want to know how pioneer plants differ from cluster plants in their labor sourcing strategies. In our regression framework, this question becomes how the pre-entry RCA* of a plant's region-industry cell affects the distance over which workers are sourced, where distance can either be the distance in geographical or in skill space. To answer it, we estimate the following equation:

$$d_{et} = \alpha_t + \gamma_i + \delta_r + \beta_{RCA} RCA^*_{irt-1} + \epsilon_{et} \tag{11}$$

$d_{et}$ is the distance between the old and the new job of an employee $e$ who starts working in a new plant in year $t$. We estimate (11) for six different types of distances, four dichotomous and two continuous ones. The dichotomous distances indicate whether or not, a worker, when entering a new plant, switches industry, occupation or region or comes out of inactivity. The continuous distances are the road distance in kilometers between the old and the new job and the skill relatedness (which can be regarded as an inverse skill distance) between the worker's old and new industry. As explained in section 4, these distances are expressed in standard deviations of the mean values of these variables for all job switchers in industry $i$ in year $t$ in the overall economy.

[12] In a later version of this text, we will plan to calculate this variable using a plant's contemporaneous employment.

[13] This effect may be due to attrition or due to changes in the occupational mixes within plants. In either case, we find that the initial differences between plants in low and high RCA locations disappears over time.

$\alpha_t$, $\gamma_i$ and $\delta_r$ are year, industry and region fixed effects. These effects absorb differences in mobility that are associated with general regional, industry and year characteristics, such as the remoteness of the region or the overall spatial clustering of an industry. The parameter of interest, $\beta_{RCA}$, is thus estimated off the variation in how far the mobility of workers in the new plants in a given region deviates from the average mobility patterns in that region, controlling for the average of this deviation across plants that enter the same industry. It represents the extent to which the pre-entry RCA* of the new plant's industry-location cell affects the distance across which workers are sourced. Negative values indicate that for workers in low RCA places (i.e., in pioneer plants) the distance to their old jobs is larger than for workers in high RCA places (i.e., in cluster plants). Note that $\beta_{RCA}$ does not tell us anything about the distance workers travel to pioneer plants *per se*, but rather how different workers in pioneer and cluster plants are in terms of mobility.

Figure 5 shows estimates for $\beta_{RCA}$ by distance type. Each distance type is negatively related to pre-entry RCA*.[14] In other words, the lower a plant's own industry's concentration in the region, the higher the likelihood that it will recruit its workers from different industries, different regions, from outside the labor force, over larger distances and from less related industries. Moreover, workers that enter new plants in low-RCA locations also more often change their occupation than those who enter plants in high-RCA locations.

Figure 5 about here

The model in section 3 proposes that the more tradable the goods are that a pioneer plant produces, the more likely it is that the plant will hire workers from outside the region. In other words, the (geographical) distance over which pioneer plants source their workers will be greater if new plants produce tradable goods. The model also predicts that pioneer plants will hire workers from outside the region if a worker really needs the right skills to be productive in the new plant. As argued in section 4, this will depend on how crucial workers are to the plant's successful operation. We submit that the higher the wage and the higher the education of workers, the more likely it is they will be crucial. Consequently, we are not just interested in how pioneer plants differ from cluster plants in their labor sourcing strategies, but also in *how these differences differ* across worker and across industry types. To investigate this, we split the sample by worker and by industry types and estimate (11) for each sub-

[14] Note that the positive effect on skill relatedness translates into a negative effect on skill distance.

sample separately. Figure 6 plots the various estimates of $\beta_{RCA}$ for each sub-sample, where the vertical axis is centered on the estimated $\beta_{RCA}$ in the whole sample.[15]

Figure 6 about here

The negative association between industry switching and pre-entry RCA* is weakest for college educated workers, for highly paid workers and for workers in non-ubiquitous occupations. Unreported t-tests indicate that these differences are significant. At the same time, these workers display the strongest negative effect of pre-entry RCA* on region switching and on road distance. This finding is consistent with the notion that pioneer plants are particularly sensitive to workers' skills when it comes to employees in core positions (i.e., college educated workers, workers in rare occupations and with a high pay), which are rather filled by recruiting workers from far away than by hiring workers with a less adequate work experience.

For workers with less crucial skills the trade-off between skill and geographical distance is the opposite. Compared to cluster plants, pioneer plants will hire these workers more often locally, but from other industries. By-and-large, the results on skill relatedness support this conclusion. Although the education level does not seem to matter much, high wage earners tend to be less affected by pre-entry RCA* levels than workers with medium wages. Somewhat unexpectedly however, the effects of pre-entry RCA* on skill relatedness are strongest for low wage earners.

The bottom part of Figure 6 shows that the importance of pre-entry RCA* differs by industry type. Plants in tradable industries, i.e., plants in manufacturing and non-ubiquitous industries, are most affected by their location's pre-entry RCA* when it comes to the geographic origins of their workers, but least when it comes to their industrial origins. To be precise, the negative relation between a location's pre-entry RCA* and the probability that a worker comes from outside the region is much stronger for manufacturing than for non-manufacturing plants and for non-ubiquitous than for ubiquitous industries. The opposite holds for the probability that workers come from different industries, switch occupations or come out of inactivity: the association with pre-entry RCA* is now *weakest* for manufacturing and non-ubiquitous industries. This suggests that plants in tradable industries indeed get experienced workers by hiring them from regions farther away when they are locally scarce, instead of hiring less experienced workers. In contrast, industries that are more likely to service local markets hire local

[15] The underlying regression outcomes and t-tests are provided in the Appendix in Tables A1 and A2.

workers, without prior industry experience. The continuous distance variables point towards the same conclusion. Manufacturing and non-ubiquitous industries display the largest pre-entry RCA* effects on road distance, but the smallest on skill relatedness.

Figures 5 and 6 show that the difference in distance over which pioneer and cluster plants source their workforces is affected both by the type of worker that is being hired and by the type of industry to which the plant belongs. However, it is possible that tradable industries employ more highly skilled workers. In this case, it is unclear to what extent the differences between tradable and non-tradable industries reflect differences in worker types or vice versa. To investigate this, we pool workers across all samples and allow the effect of pre-entry RCA* to differ by worker and industry type by adding appropriate interaction terms. The regression equation now becomes:

$$d_{et} = \alpha_t + \gamma_i + \delta_r + \sum_{s\in S} \beta_s s(e) + \beta_T T(i) +$$

$$\beta_{RCA} RCA^*_{irt-1} + \sum_{s\in S} \beta_{s\times RCA} s(e) RCA^*_{irt-1} + \beta_{T\times RCA} RCA^*_{irt-1} T(i) + \epsilon_{et} \quad (12)$$

$\alpha_t$, $\gamma_i$ and $\delta_r$ are again year, industry and region fixed effects. $s(e)$ represents a dummy group for the skill type of worker $e$ and $T(i)$ is a dummy variable that equals 1 if $e$ works in a tradable industry. Skill types are defined by workers' wage (low, medium and high wages) and education levels (low, medium, and college education or missing if no education information is provided), where the reference group is composed of workers with low wages, respectively with a low education level.

As an example, Table 3 provides the estimates for road distance of four different variants of equation (12), where worker type is based on education levels and the industry type by the manufacturing/non-manufacturing split. Models (1) and (2) contain worker and industry type dummies (reported in the upper panel of the table), region, industry and year fixed effects and the pre-entry RCA* of the plant's industry-location cell.[16] College educated workers travel longer distances than workers with medium levels of education, who travel farther than workers with a low level of education. Column (2) adds industry-worker type interactions and shows that geographical mobility is highest for college educated workers in manufacturing industries. In both models (1) and (2), the effect of pre-entry RCA* is negative, but modest. A one unit increase in pre-entry RCA*, which corresponds roughly to the difference in RCA* between pioneer plants and the average plant in our sample, is associated with a 3.8% standard

[16] Note that the industry effects absorb the intercept of the manufacturing dummy.

deviations (about 6 km) shorter distance. Column (3) adds the interactions of pre-entry RCA* with worker and industry type dummies. Column (4) also adds three-way interactions between industry type, worker type and pre-entry RCA*. The lower panel of Table 3 sums estimates in the middle panel to provide total pre-entry RCA* effects by worker-industry type combination. Pre-entry RCA* effects for a given worker-industry type combination do not differ much between models (3) and (4), nor do the three-way interactions add explanatory power. Henceforth, we therefore focus on the coefficients in model (3).

Table 3 about here

The intercepts in the upper panel do not change much when including pre-entry RCA* interaction terms. However, the middle panel shows that the negative effect of pre-entry RCA* can be attributed completely to college educated and manufacturing workers. In other words, pioneer plants do not differ from cluster plants in the distance over which they hire workers with low or medium levels of educations or if they belong to non-manufacturing industries. In a typical manufacturing plant, workers are sourced from about 0.07 to 0.10 standard deviations (11 to 15 km) farther away for a one unit increase in pre-entry RCA*. For college educated workers, a one unit increase in pre-entry RCA* adds another 28 km.

Because we define worker types and industry types in different ways, we get different estimates for each pairing of definitions. Tables 4 and 5 summarize the results of model (3) for each of these pairings. The distances we investigate are region switching, industry switching, road distance and skill relatedness. Table 4 shows the intercepts of the upper panel in Table 3 and Table 5 shows the interaction effects that are summarized in Table 3's middle panel.

First, we note that the definition of tradability used in the interaction terms reported in Table 5 does not affect the intercepts in Table 4. Table 4 furthermore shows that the likelihood that workers come from outside the new plant's region (as indicated by the intercepts in equation (12)) falls monotonically with the workers' education and wage levels. These effects are quantitatively large. For instance, the probability that highly educated workers come from outside the region is 18.9 percentage points higher than the corresponding probability for workers with low education levels. Road distance shows the same monotonic decline with education and wage levels. However, for industry switching and skill relatedness, outcomes are more ambiguous, with some specifications showing low-skill workers to be more mobile across industries than medium skill workers, while other specifications show the reverse.

Our main interest is in the pre-entry RCA* parameter estimates in Table 5. The negative main effects of pre-entry RCA* indicate that workers in pioneer plants (low RCA places) are more likely than workers in cluster plants (high RCA places) to switch regions and industries. These workers also come more often from less related industries. The negative interaction terms show that, in terms of region switching, differences between pioneer plants and cluster plants are most pronounced for highly skilled workers. A one unit increase in pre-entry RCA* leads to an 8.0% to 11.8% higher probability that the worker came from outside the new plant's region. For low-skill workers, this increase is just the main effect (4.2%-5.0%). In contrast, the negative pre-entry RCA* effect on industry switching becomes weaker the higher the skills of a worker gets, indicating a narrowing of the difference between pioneer and cluster plants with rising skill levels. The pre-entry RCA* effect on road distance behaves much like the effect on region switching: it is negative and strongest for high-skill workers, than for medium-skill workers and absent for low-skill workers. The pre-entry RCA* effect on skill relatedness is positive (i.e., workers in pioneer plants come from less related industries than workers in cluster plants).

Turning to the industry type interactions, we find that the tradability of an industry also modifies pre-entry RCA* effects, although the evidence here is weaker. For instance, the difference in industry switching between pioneer and cluster plants narrows if plants belong to non-ubiquitous (rare) industries. A similar narrowing is found for skill relatedness. Moreover, although there is no evidence for differential effects on region switching, pioneer plants in both manufacturing and in non-ubiquitous industries source their workers over longer distances than cluster plants do. Hence, the differences in pre-entry RCA* effects by industry in Figure 6 are not all just due to differences in workforce composition between plants in tradable and non-tradable industries. Pioneer plants seem to balance the trade-off between hiring experienced workers and hiring workers locally differently if they produce tradable goods than if they produce non-tradable goods. In tradable industries, pioneer plants are more willing to hire workers from far away, while compromising less on worker's skills, whereas the reverse holds for pioneer plants in non-tradable industries.

## 6. Conclusions

We studied whether plants that are regional pioneers in their industries differ from other plants in from where they recruit their workforce. We find that most plants enter regions in which their industry is already well-represented. Moreover, plants in such high RCA locations tend to have occupational mixes that are closer to the industry average, but this difference diminishes as plants grow older.

Outside the main clusters of their industries, plants cannot easily hire local workers with experience in the plants' industries, which is reflected in the origins of their workforces. Plants in low-RCA locations are more likely to recruit workers that previously worked in other regions that are farther away and from other industries that are less related than plants in high-RCA locations do. Moreover, plants in low RCA places also more often hire workers from outside the labor market or who worked in a different occupation from the one they are hired for at the new plant. The differences between plants in low and high RCA places vary by worker and by industry type. Pioneer plant's tendency to hire workers from far away and not from less related industries is particularly strong for workers that are highly educated or highly paid. We posit that such high-skill workers are often key to the successful operation of a plant, which is why plants cannot afford to compromise on their skills.

Similar observations hold for tradable, i.e., manufacturing and non-ubiquitous, industries. In tradable industries, the RCA of a location has a relatively large impact on the distance over which workers are recruited, but it affects the experience mismatch that plants accept much less. Our interpretation of these findings, in line with our model, is that manufacturing and non-ubiquitous industries have less pricing power and hence are less able to take the risk of a poorly matched labor force. Therefore, they are less willing to hire workers from more distant industries and instead prefer to hire in more distant locations.

Our study raises a number of puzzles for follow-up research. For instance, an interesting question is how plants outside clusters manage to hire workers over larger distances. Does this require higher wages for workers of the same level of quality? Or are the costs of living lower in these new locations? Moreover, one could study to what extent worker and industry characteristics interact: which employees are critical in which industries? Finally, one could try to assess which regions, given their location and industry mix would make suitable candidates to attract a certain industry and for which regions this industry would indeed be very hard to attract. One important caveat is that we only look at plants that manage to survive the first three years of their existence. Adding plants that fail within that time span would allow us to see to what extent failing to find the right workers forces plants to shut down.

Marshall expected that with the invention of the telegraph and the introduction of railways, locating in places with a specialized labor force might become less important. Our findings show that plants that locate outside the major clusters of their industries also today face starker choices in assembling their workforce. To some extent, they opt for hiring workers with seemingly less relevant work experience. However, the more critical the skills of a worker and the more a plant is exposed to competition from other regions, the more often pioneer plants resort to recruit their workers from farther away. In the context of an advanced economy with a well-functioning infrastructure and a highly trained labor force, enlarging the area in which to search for adequately skilled workers is certainly feasible. However, even in Germany, new plants inside clusters vastly outnumber those outside these clusters, suggesting that the feasibility of entrepreneurial venture still depends on the local availability of the right workers. In the context of developing economies, searching for and recruiting adequately skilled workers over large distances is arguably much more costly. In fact, the right workers may not even exist inside the country. Under such conditions pioneering a new industry may be all but impossible. Even today, being in a location that provides access to the right workers still seems crucial.

## 7. References


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

Almeida P., Kogut B. (1999). Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science*, 45(7): 905-991.

B. Balassa, Comparative advantage in manufactured goods: a reappraisal, Review of Economics and Statistics 68, 315 (1986).

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

Bonache J, Brewster C. 2001. Knowledge Transfer and the Management of Expatriation. Thunderbird International Business Review, 43(1):145–168

Boschma R.A., Frenken K. (2009) Technological relatedness and regional branching, in: Bathelt H., Feldman M.P., Kogler D.F.(eds.), *Dynamic Geographies of Knowledge Creation and Innovation*, Routledge, Taylor and Francis, forthcoming.

Boschma, R., Minondo, A., Navarro, M. (2013), The emergence of new industries at the regional level in Spain. A proximity approach based on product-relatedness, Economic Geography, 89 (1), 29-51.

Breschi S, Lissoni F. 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. Journal of Economic Geography, 9(4): 439-468

Buenstorf G, Klepper S. 2009. Heritage and Agglomeration: The Akron Tyre Cluster Revisited. Economic Journal, 119(537):705-733.

Cooke P.H., and Morgan K., 1998. *The Associational Economy. Firms, Regions, and Innovation.* Oxford: Oxford University Press.

Collins H, Pinch T. 1993. *The Golem: What everyone should know about science*. New York: Cambridge University Press.

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

Dahl MS, Sorenson O. 2012. Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures. Management Science 58(6): 1059–1071.

Dahl MS, Sorenson O. 2012. The embedded entrepreneur European Management Review 6: 172-181.

Duranton G., and Puga D., 2004. Micro-Foundations of Urban Agglomeration Economies. In: *Handbook of Urban and Regional Economics* (4th edition). Henderson J., and Thisse J.-F. (eds.). Amsterdam: Elsevier.

Fosfuria A, Motta M, Rønde Th. 2001. Foreign direct investment and spillovers through workers' mobility. Journal of International Economics, 53: 205–222

Glaeser E.L., Kallal H.D., Scheinkman J.A., and Schleifer A., 1992. Growth in Cities. *Journal of Political Economy* (100:6) pp. 1126-1152.

Hausmann, R., Klinger B. (2007) The structure of the product space and the evolution of comparative advantage. CID working paper no. 146, Center for International Development, Harvard University, Cambridge.

Head K, Ries J,Swenson D. 1995. Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States. Journal of International Economics, 38: 223-247.

Hébert L, Very Ph, Beamish PW. 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 W. C. Strange (1990), "Agglomeration Economies and Matching in a System of Cities." Regional Science and Urban Economics, 20: 189-212.

Henderson, J.V., Kuncoro, A., Turner, M. (1995) Industrial development in cities, *Journal of Political Economy*, 103:1067-1085.

Hethey, T.; Schmieder, J. F. (2010): Using worker flows in the analysis of establishment turnover. Evidence from German administrative data. (FDZ Methodenreport, 06/2010), Nürnberg.

Hidalgo C.A., Klinger B., Barabási A.-L., and Hausmann R., 2007. The Product Space Conditions the Development of Nations. *Science* (317) pp. 482-487.

Keller W. 2004. International Technology Diffusion. Journal of Economic Literature, 17: 752–782.

Long LH. 1973. Migration Differentials by Education and Occupation: Trends and Variations.Demography, 10(2): 243-258

MacKenzie, D., G. Spinardi (1995). Tacit knowledge, weapons design and the uninvention of nuclear weapons. *American Journal of Sociology*, 101(1): 44-99.

Marshall A. 1920. *Principles of Economics*. 8th edition. London: Macmillan and Co. Original edition 1890.

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

Mendenhall ME, Dunbar E, Oddou GR. 1987. Expatriate selection, training and career-pathing: A review and critique. Human Resource Management, 26(3): 331-445.

Miller EL. 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 FMH, Henning M. (2013). Skill-relatedness and firm diversification. Strategic Management Journal, 34(3): 297-316

Neffke, F., M. Henning and R. Boschma (2011), How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, Economic Geography, 87 (3): 237-265.

Polanyi M., 1967. *The Tacit Dimension.* New York: Doubleday.

Porter M. 2003. The economic performance of regions. *Regional Studies* **37**(6-7): 549-578.

Porter M.E., 1998. Clusters and the New Economics of Competition*. Harvard Business Review*. November-December.

Rosenthal S.S., and Strange W.C., 2003. Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics* (85:2) pp. 377-393.

Timmermans B. 2010. Old Wine in New Bottles: The Effect of Previous Co-Worker Experience on the Survival of New Firms. mimeo. DRUID Working Papers 10-17.

# 8. Tables and Figures

Table 1: Number of new firms and their employees by sample

| NUMBER OF NEW PLANTS | | Number of new plants | As % of total |
|---|---|---|---|
| **all industries** | | 19,710 | 100.0% |
| **manufacturing** | **manufacturing** | 1,613 | 8.2% |
| | **non-manufacturing** | 18,097 | 91.8% |
| | | | |
| **ubiquity** | **non-ubiquitous** | 1,641 | 8.3% |
| | **ubiquitous** | 18,069 | 91.7% |
| **NUMBER OF WORKERS** | | **Number of workers** | **As % of total** |
| **all workers, all industries** | | 1,712,085 | 100.0% |
| ***worker types*** | | | |
| **wages** | **low** | 428,233 | 25.0% |
| | **mid** | 855,740 | 50.0% |
| | **high** | 428,112 | 25.0% |
| | | | |
| **education** | **low** | 318,180 | 18.6% |
| | **mid** | 815,963 | 47.7% |
| | **college** | 76,644 | 4.5% |
| | **not available** | 501,298 | 29.3% |
| **occupations** | **non-ubiquitous** | 257,083 | 15.0% |
| | **ubiquitous** | 1,455,002 | 85.0% |
| | | | |
| ***industry types*** | | | |
| **manufacturing** | **manufacturing** | 108,728 | 6.4% |
| | **non-manufacturing** | 1,603,357 | 93.6% |
| | | | |
| **ubiquity** | **non-ubiquitous** | 127,792 | 7.5% |
| | **ubiquitous** | 1,584,293 | 92.5% |

The table is based on all selected new plants and all employees that enter any these plants between 1999 and 2008. RCA* is reported once as the average RCA* across all industry-region combinations in the German economy and once as the average for industry-region combinations weighted by their new plant employment.

Table 2: Descriptive statistics for variables in study

| Variable | Mean | St. dev. |
|---|---|---|
| **RCA*** | | |
| **across all regXind cells** | -0.412 | 0.542 |
| **weighted by new plant emp.** | -0.005 | 0.321 |
| **SWITCHING** | | |
| **from inactivity** | 0.622 | 0.485 |
| **from other firms** | 0.378 | 0.235 |
| **industry switching** | 0.748 | 0.434 |
| **region switching** | 0.430 | 0.495 |
| **occupation switching** | 0.602 | 0.490 |
| | | |
| **CONTINUOUS DISTANCES** | | |
| **road distance (km)** | 97.9 | 153.3 |
| **skill relatedness** | 0.490 | 0.527 |

The table is based on all employees that enter any of the selected new plants between 1999 and 2008. RCA* is reported once as the average RCA* across all industry-region combinations in the German economy and once as the average for industry-region combinations weighted by their new plant employment. The statistics for industry, region and occupation switchers are based on workers that were employed in another firm in the previous year.

Table 3: Regression of road distance on pre-entry RCA*

| *Dep. Var.: Road distance (st. dev.s)* | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| **INTERCEPTS** | ***education*** | | | | |
| | col | 0.415*** | 0.393*** | 0.420*** | 0.401*** |
| | | (0.022) | (0.023) | (0.022) | (0.024) |
| | mid | 0.149*** | 0.150*** | 0.149*** | 0.150*** |
| | | (0.012) | (0.012) | (0.012) | (0.012) |
| | miss | 0.162*** | 0.161*** | 0.161*** | 0.160*** |
| | | (0.015) | (0.015) | (0.015) | (0.015) |
| | ***industry X education*** | | | | |
| | man X col | | 0.226*** | | 0.182*** |
| | | | (0.057) | | (0.057) |
| | man X mid | | 0.003 | | -0.009 |
| | | | (0.036) | | (0.037) |
| | man X miss | | 0.017 | | -0.002 |
| | | | (0.047) | | (0.045) |
| **PRE-ENTRY RCA*** | ***main effect*** | | | | |
| | pre-entry RCA* | -0.038** | -0.038** | 0.013 | -0.023 |
| | | (0.019) | (0.019) | (0.045) | (0.049) |
| | ***education*** | | | | |
| | col X RCA* | | | -0.185*** | -0.121** |
| | | | | (0.053) | (0.060) |
| | mid X RCA* | | | -0.031 | 0.002 |
| | | | | (0.039) | (0.044) |
| | miss X RCA* | | | -0.024 | 0.018 |
| | | | | (0.046) | (0.051) |
| | ***industry*** | | | | |
| | man X RCA* | | | -0.084** | 0.127 |
| | | | | (0.038) | (0.100) |
| | ***education X industry*** | | | | |
| | col X man X RCA* | | | | -0.248* |
| | | | | | (0.136) |
| | mid X man X RCA* | | | | -0.215** |
| | | | | | (0.100) |
| | miss X man X RCA* | | | | -0.269** |
| | | | | | (0.117) |
| **FIXED EFFECTS** | industry | yes | yes | yes | yes |
| | region | yes | yes | yes | yes |
| | year | yes | yes | yes | yes |
| **TOTAL RCA* EFFECTS** | col & man | -0.038** | -0.038** | -0.257*** | -0.265 |
| | | (0.019) | (0.019) | (0.079) | (0.186) |
| | col & non-man | -0.038** | -0.038** | -0.173*** | -0.143* |
| | | (0.019) | (0.019) | (0.069) | (0.077) |
| | mid & man | -0.038** | -0.038** | -0.103 | -0.109 |
| | | (0.019) | (0.019) | (0.070) | (0.156) |
| | mid & non-man | -0.038** | -0.038** | -0.018 | -0.021 |
| | | (0.019) | (0.019) | (0.059) | (0.066) |
| | low & man | -0.038** | -0.038** | -0.072 | 0.104 |
| | | (0.019) | (0.019) | (0.058) | (0.111) |
| | low & non-man | -0.038** | -0.038** | 0.013 | -0.023 |
| | | (0.019) | (0.019) | (0.045) | (0.049) |
| **R-squared** | **R-squared** | 0.015 | 0.0197 | 0.0197 | 0.0199 |

Regression of road distance between old and new job of workers entering new plants in the first five years of the plant's existence. Industries are tradable if they belong to the manufacturing sector (man=1). Workers are grouped by education levels (col: college education, mid: medium level education, miss: education information is missing, omitted: low level education). Lower panel contains total effects by worker-industry combination. p-values: ***: p<.01; **: p<.05; *: p<.10. Standard errors are clustered by plant (N=19,673).

Table 4: Intercepts by worker and industry type

| DISTANCE | CLASSIFICATION | | WORKER TYPE | |
|---|---|---|---|---|
| | tradability | skill | high | medium |
| **Region switching** | **manufacturing** | **education** | 0.189*** | 0.079*** |
| | | | (0.008) | (0.005) |
| | | **wages** | 0.183*** | 0.062*** |
| | | | (0.006) | (0.004) |
| | **non-ubiquitous** | **education** | 0.189*** | 0.079*** |
| | | | (0.008) | (0.005) |
| | | **wages** | 0.183*** | 0.062*** |
| | | | (0.006) | (0.004) |
| **Industry switching** | **manufacturing** | **education** | -0.001 | 0.022*** |
| | | | (0.006) | (0.004) |
| | | **wages** | -0.045*** | 0.013*** |
| | | | (0.005) | (0.003) |
| | **non-ubiquitous** | **education** | -0.001 | 0.022*** |
| | | | (0.006) | (0.004) |
| | | **wages** | -0.045*** | 0.013*** |
| | | | (0.005) | (0.003) |
| **Road distance** | **manufacturing** | **education** | 0.420*** | 0.149*** |
| | | | (0.022) | (0.012) |
| | | **wages** | 0.374*** | 0.111*** |
| | | | (0.014) | (0.010) |
| | **non-ubiquitous** | **education** | 0.420*** | 0.149*** |
| | | | (0.022) | (0.012) |
| | | **wages** | 0.374*** | 0.111*** |
| | | | (0.014) | (0.010) |
| **Skill relatedness** | **manufacturing** | **education** | 0.057*** | -0.049*** |
| | | | (0.017) | (0.010) |
| | | **wages** | 0.267*** | 0.064*** |
| | | | (0.013) | (0.008) |
| | **non-ubiquitous** | **education** | 0.057*** | -0.049*** |
| | | | (0.017) | (0.010) |
| | | **wages** | 0.266*** | 0.064*** |
| | | | (0.013) | (0.008) |

Intercepts in regressions on distances between old and new job of workers entering new plants in the first five years of the plant's existence as specified in model (3) of Table 3. Column 2 depicts the definition of tradability: industries are defined as tradable if they belong to (a) the manufacturing sector or (b) are non-ubiquitous (*rare*). Column 3 shows whether workers' skill levels are defined using education or wage levels. In Columns 3 and 4, *high* stands for college (*Fachhochschule* or *Hochschule*) educated (education rows) or for wages above the $75^{th}$ percentile (wage rows), *medium* for medium level education (*Abitur* or high school plus vocational training) or for wages between the $25^{th}$ and $75^{th}$ percentile, low education and wage levels are the omitted category. p-values: ***: p<.01; **: p<.05; *: p<.10. Standard errors are clustered by plant (N=19,673).

Table 5: Pre-entry RCA* effects by worker and industry type

| DISTANCE | CLASSIFICATION | | MAIN EFFECT | WORKER TYPE | | INDUSTRY TYPE |
|---|---|---|---|---|---|---|
| | tradability | skill | | high | medium | tradable |
| **Region switching** | **manufacturing** | **education** | -0.048*** | -0.069*** | -0.001 | -0.017 |
| | | | (0.018) | (0.022) | (0.017) | (0.017) |
| | | **wages** | -0.042*** | -0.038** | 0.017 | -0.021 |
| | | | (0.013) | (0.015) | (0.013) | (0.016) |
| | **non-ubiquitous** | **education** | -0.050*** | -0.068*** | -0.001 | -0.004 |
| | | | (0.018) | (0.022) | (0.017) | (0.014) |
| | | **wages** | -0.045*** | -0.040*** | 0.016 | 0.006 |
| | | | (0.013) | (0.015) | (0.013) | (0.014) |
| **Industry switching** | **manufacturing** | **education** | -0.098*** | 0.066*** | 0.031*** | 0.012 |
| | | | (0.012) | (0.016) | (0.011) | (0.012) |
| | | **wages** | -0.106*** | 0.056*** | 0.030*** | 0.010 |
| | | | (0.009) | (0.011) | (0.009) | (0.012) |
| | **non-ubiquitous** | **education** | -0.106*** | 0.057*** | 0.027** | 0.062*** |
| | | | (0.012) | (0.016) | (0.011) | (0.010) |
| | | **wages** | -0.113*** | 0.048*** | 0.028*** | 0.058*** |
| | | | (0.009) | (0.011) | (0.009) | (0.010) |
| **Road distance** | **manufacturing** | **education** | 0.013 | -0.185*** | -0.031 | -0.084** |
| | | | (0.045) | (0.053) | (0.039) | (0.038) |
| | | **wages** | -0.002 | -0.079** | 0.030 | -0.092** |
| | | | (0.032) | (0.035) | (0.029) | (0.037) |
| | **non-ubiquitous** | **education** | -0.001 | -0.185*** | -0.031 | 0.004 |
| | | | (0.044) | (0.052) | (0.039) | (0.037) |
| | | **wages** | -0.015 | -0.088** | 0.023 | 0.022 |
| | | | (0.032) | (0.036) | (0.029) | (0.037) |
| **Skill relatedness** | **manufacturing** | **education** | 0.145*** | 0.004 | -0.018 | -0.016 |
| | | | (0.032) | (0.043) | (0.031) | (0.035) |
| | | **wages** | 0.184*** | -0.044 | -0.055** | -0.015 |
| | | | (0.026) | (0.033) | (0.026) | (0.034) |
| | **non-ubiquitous** | **education** | 0.163*** | 0.022 | -0.010 | -0.134*** |
| | | | (0.031) | (0.043) | (0.031) | (0.038) |
| | | **wages** | 0.201*** | -0.026 | -0.050* | -0.126*** |
| | | | (0.026) | (0.032) | (0.026) | (0.038) |

Main and interaction effects of pre-entry RCA* in regressions on distances between old and new job of workers entering new plants in the first five years of the plant's existence as specified in model (3) of Table 3. Column 2 depicts the definition of tradability: industries are defined as tradable if they belong to (a) the manufacturing sector or (b) are non-ubiquitous (*rare*). Column 3 shows whether workers' skill levels are defined using education or wage levels. In Columns 3 and 4, *high* stands for college (*Fachhochschule* or *Hochschule*) educated (education rows) or for wages above the 75$^{th}$ percentile (wage rows), *medium* for medium level education (*Abitur* or high school plus vocational training) or for wages between the 25$^{th}$ and 75$^{th}$ percentile, low education and wage levels are the omitted category. p-values: ***: p<.01; **: p<.05; *: p<.10. Standard errors are clustered by plant (N=19,673).

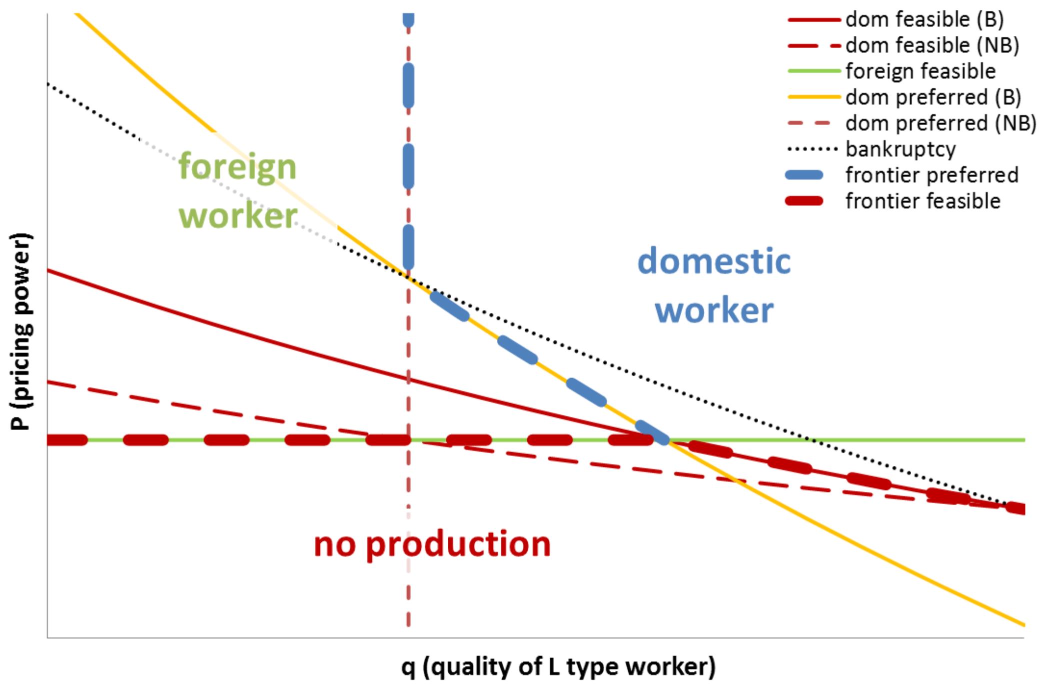


***Figure 1: Production with domestic or with foreign workers***

The lines show the various inequalities derived above. We use the following abbreviations: dom: domestic worker, B: with bankruptcy risk, NB without bankruptcy risk. Bankruptcy is the line below which a plant faces a bankruptcy risk. The dotted bold lines divide the graph into three zones, one where production cannot take place at all, one in which production with foreign workers is preferred and one in which production with domestic workers is preferred.

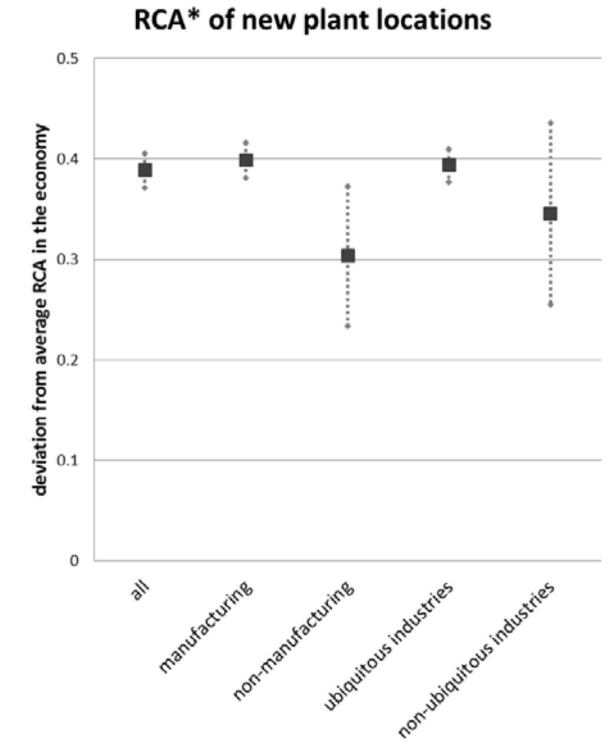


***Figure 2: RCA* of industry-region combination that attract at least one new plant***

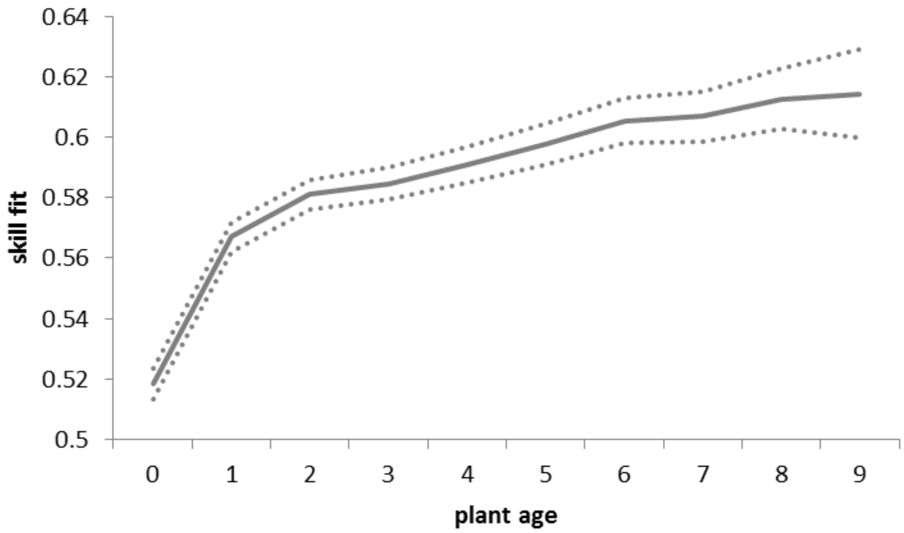


*Figure 3: Evolution of skill fit with plant age*

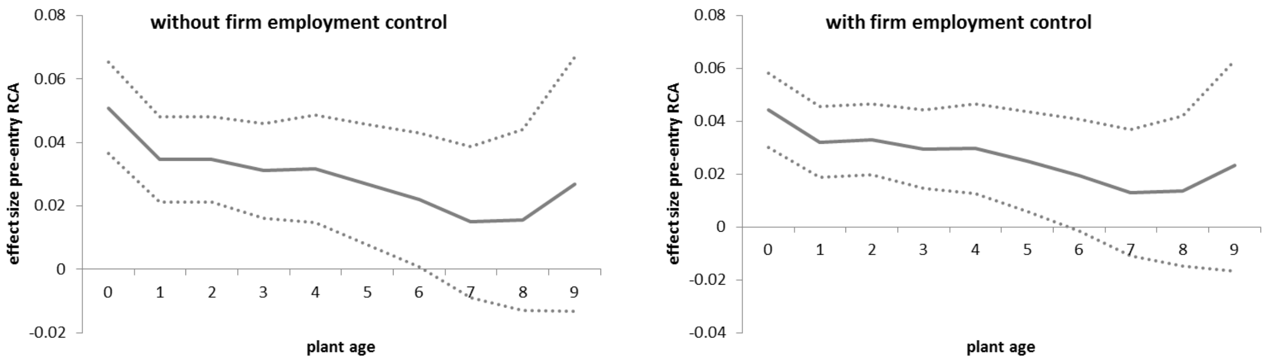


***Figure 4: Evolution of the effect of pre-entry RCA on skill fit with plant age***

| **Variable** | **parameter estimates (s.e.)** |
|---|---|
| industry switch | -0.066** |
| | (0.002) |
| region switch | -0.051** |
| | (0.002) |
| occupation switch | -0.019** |
| | (0.002) |
| from inactivity | -0.025** |
| | (0.001) |
| distance(km) | -0.031** |
| | (0.005) |
| skill relatedness | 0.148** |
| | (0.005) |

p-values: **: .01, *: .05.

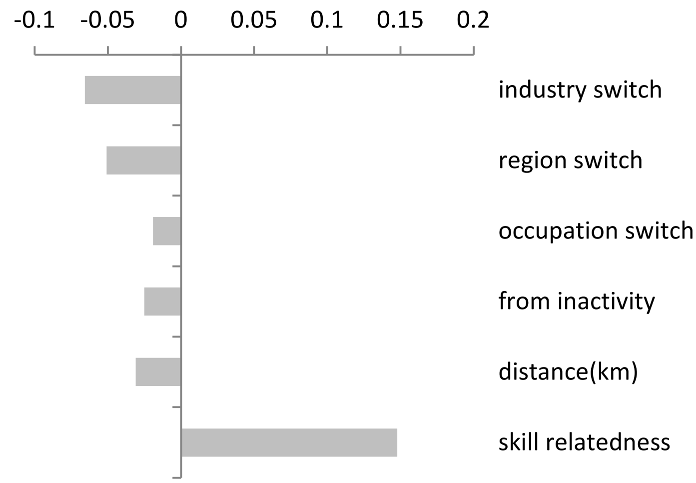


***Figure 5: Effects of pre-entry RCA* on distance variables***

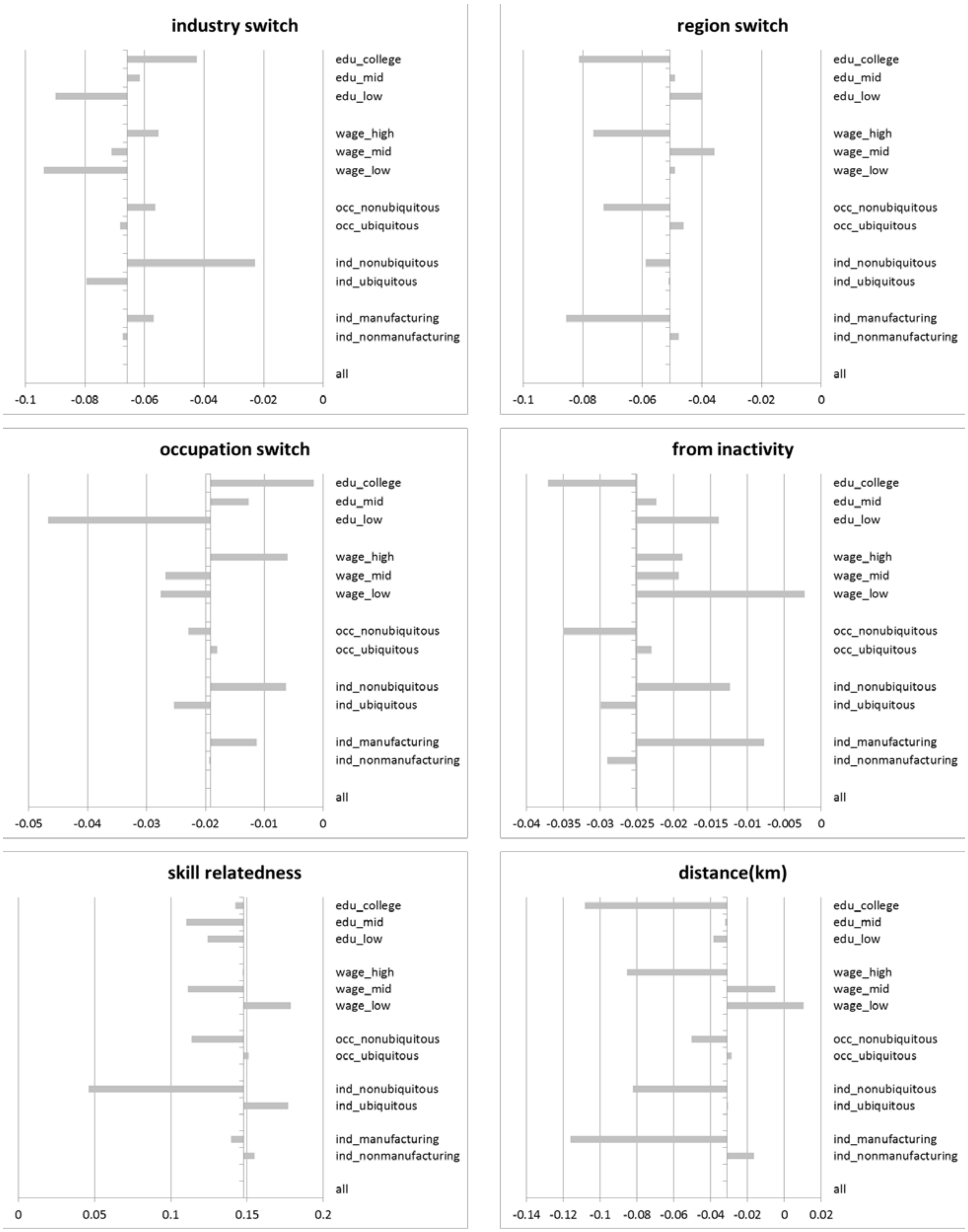


***Figure 6: Effect of pre-entry RCA* on distance variables by worker and industry type***

## Appendix

Table A1: the relation between RCA* and workers' origins by worker type

| | occupation | | wage level | | | education | | |
|---|---|---|---|---|---|---|---|---|
| **estimates** | **(1) ubiquitous** | **(2) non-ubiquitous** | **(3) low** | **(4) mid** | **(5) high** | **(6) low** | **(7) mid** | **(8) college** |
| **industry switch** | -0.068** | -0.056** | -0.094** | -0.071** | -0.055** | -0.090** | -0.062** | -0.042** |
| | (0.002) | (0.004) | (0.006) | (0.003) | (0.003) | (0.007) | (0.003) | (0.006) |
| **region switch** | -0.046** | -0.073** | -0.049** | -0.036** | -0.076** | -0.040** | -0.049** | -0.081** |
| | (0.002) | (0.005) | (0.007) | (0.003) | (0.003) | (0.007) | (0.003) | (0.008) |
| **occupation switch** | -0.018** | -0.023** | -0.028** | -0.027** | -0.006 | -0.047** | -0.013** | -0.002 |
| | (0.002) | (0.004) | (0.006) | (0.003) | (0.003) | (0.007) | (0.003) | (0.008) |
| **from inactivity** | -0.023** | -0.035** | -0.002 | -0.019** | -0.019** | -0.014** | -0.022** | -0.037** |
| | (0.001) | (0.003) | (0.003) | (0.002) | (0.002) | (0.004) | (0.002) | (0.006) |
| **distance(km)** | -0.029** | -0.050** | 0.010 | -0.005 | -0.085** | -0.039** | -0.032** | -0.108** |
| | (0.005) | (0.012) | (0.013) | (0.007) | (0.008) | (0.013) | (0.007) | (0.023) |
| **skill relatedness** | 0.151** | 0.113** | 0.179** | 0.111** | 0.147** | 0.124** | 0.110** | 0.142** |
| | (0.006) | (0.011) | (0.017) | (0.008) | (0.007) | (0.018) | (0.007) | (0.018) |
| **t-tests** | | **(2)-(1)** | **(5)-(3)** | **(5)-(4)** | **(4)-(3)** | **(8)-(6)** | **(8)-(7)** | **(7)-(6)** |
| **industry switch** | | 2.66 | 5.68 | 4.11 | 3.29 | 4.93 | 2.77 | 3.72 |
| **region switch** | | -4.91 | -3.72 | -8.82 | 1.78 | -3.77 | -3.70 | -1.17 |
| **occupation switch** | | -0.98 | 2.98 | 4.53 | 0.11 | 4.09 | 1.26 | 4.29 |
| **from inactivity** | | -3.49 | -4.71 | 0.17 | -5.33 | -3.28 | -2.37 | -1.93 |
| **distance(km)** | | -1.67 | -6.30 | -7.29 | -1.05 | -2.65 | -3.23 | 0.44 |
| **skill relatedness** | | -3.04 | -1.73 | 3.48 | -3.63 | 0.70 | 1.65 | -0.72 |

**: .01, *: .05. t-tests for parameter estimates in pairs of columns are calculated by dividing the difference in estimates by the square root of the sum of the squared standard errors, assuming that the covariances are either positive (implicating t-tests are conservative) or negligible.

Table A2: the relation between RCA* and workers' origins by industry type

| | Industries by sector | | Industries by ubiquity | |
|---|---|---|---|---|
| **estimates** | **(1) non-manufacturing** | **(2) manufacturing** | **(3) ubiquitous** | **(4) non-ubiquitous** |
| **industry switch** | -0.067** | -0.057** | -0.079** | -0.023** |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| **region switch** | -0.048** | -0.086** | -0.051** | -0.059** |
| | (0.002) | (0.006) | (0.002) | (0.005) |
| **occupation switch** | -0.019** | -0.011* | -0.025** | -0.006 |
| | (0.002) | (0.005) | (0.002) | (0.005) |
| **from inactivity** | -0.029** | -0.008 | -0.030** | -0.012** |
| | (0.001) | (0.004) | (0.002) | (0.003) |
| **distance(km)** | -0.017** | -0.116** | -0.031** | -0.082** |
| | (0.005) | (0.014) | (0.006) | (0.013) |
| **skill relatedness** | 0.155** | 0.139** | 0.177** | 0.046** |
| | (0.005) | (0.013) | (0.006) | (0.011) |
| **t-tests** | | **(2)-(1)** | | **(4)-(3)** |
| **industry switch** | | 2.60 | | 16.49 |
| **region switch** | | -5.85 | | -1.41 |
| **occupation switch** | | 1.37 | | 3.67 |
| **from inactivity** | | 4.83 | | 5.22 |
| **distance(km)** | | -6.62 | | -3.71 |
| **skill relatedness** | | -1.07 | | -10.58 |

p-values: **: .01, *: .05. t-tests for parameter estimates in pairs of columns are calculated by dividing the difference in estimates by the square root of the sum of the squared standard errors, assuming that the covariances are either positive (implicating t-tests are conservative) or negligible.