Inter-industry labor flows

Version: January 2017

Authors
Frank Neffke¹, Anne Otto², Antje Weyh²

¹ Corresponding author, Harvard Kennedy School, Center for International Development,
² Institute for Employment Research (IAB)

Keywords: labor mobility, relatedness, skills, regional growth, Germany, human capital specificity

Abstract
Using German social security data, we study inter-industry labor mobility to assess how industry-specific human capital is and to determine which industries have similar human capital requirements. We find that inter-industry labor flows are highly concentrated in just a handful of industry pairs. Consequently, labor flows connect industries in a sparse network. We interpret this network as an expression of industries’ similarities in human capital requirements, or skill relatedness. This skill-relatedness network is stable over time, similar for different types of workers and independent of whether workers switch jobs locally or over larger distances. Moreover, when we use skill-relatedness as a predictive tool, it predicts regional diversification and industrial development better than colocation or value chain linkages among industries. To facilitate future research, we make detailed inter-industry relatedness matrices available in an online appendix.

Acknowledgments: We thank David Autor, Alex Coad, Koen Frenken, Andres Gomez, Ricardo Hausmann, César Hidalgo, Paul Novosad, Stuart Russell, Muhammed Yildirim and the participants of the Workshop on Industry relatedness and Regional Qualitative Change in Aalborg for valuable comments. Frank Neffke is grateful to the MasterCard Center for Inclusive Growth & Financial Inclusion for their financial support.
1. Introduction

Labor mobility plays an important role in economics. On the one hand, industry-specific technology or demand shocks often necessitate a transfer of productive capacity, and thus of workers, from shrinking to growing industries. On the other hand, labor mobility diffuses knowhow across firms, industries and locations, and is therefore important in organizational learning (March, 1991; Simon, 1991) and regional and national growth (Saxenian, 2006). Unsurprisingly, therefore, labor mobility has received much scholarly attention from both labor economists and innovation economists. However, one aspect of labor mobility has hitherto been neglected, namely, the mobility of workers across industry boundaries. This is surprising, given that any constraints to such mobility will limit both, the reallocation of labor, and the diffusion of knowledge. As a consequence, the industry structure of labor flows is still poorly understood.

In this paper, we contribute to our understanding of inter-industry labor flows, showing that they exhibit strong regularities. We summarize these regularities in a set of stylized facts that are organized around three related topics: (1) the expression of human capital specificities in the structure labor flows, (2) the use of the labor-flow-structures as measures of inter-industry relatedness and (3) the way in which constraints on inter-industry labor-flows affect diversification and labor reallocation in local economies. In particular, we ask a number of interrelated questions: Do labor flows concentrate in relatively few industry pairs? How stable is the network of inter-industry labor flows? Is this network general or specific to an occupation? Does the sparseness of the inter-industry labor-flow-network condition a region’s growth path? And, finally, does this sparseness constrain a local economy’s capacity to reallocate labor from contracting to expanding industries?

These questions complement a vast literature on general labor flows and job switching. For instance, labor economists have extensively studied job-switching rates (or, their complement, employment durations) and how they depend on business cycles, industry and worker characteristics.
Recent work in this tradition finds that workers often change jobs across industries that belong to completely different sectors (Parrado et al., 2007; Kambourov and Manovskii, 2008). This finding may lead to the conclusion that human capital has no strong industry-specific component. However, because this research fails to take into consideration which industries exchange workers, it implicitly assumes that all industries are equidistant from one another in terms of human capital requirements. We show that a closer analysis of the network structure of inter-industry labor flows casts doubts on this conclusion. These analyses are collected in a first set of stylized facts that describe the general degree of structure observed in inter-industry labor flows.

A different group of scholars at the intersection of innovation economics and economic geography have studied the role of labor flows as conduits of knowledge diffusion, typically focusing on the mobility of a subset of highly skilled workers, such as inventors. However, although the spatial limits to mobility are central in much of this research (Breschi and Lissoni, 2005; Agrawal et al., 2006; Casper, 2007), also here the question of whether there are inter-industry constraints to labor mobility has typically been neglected.

A second debate to which our work relates takes place in the literature on inter-industry relatedness. In spite of this relative neglect of inter-industry labor flows, an increasing number of papers has turned to such flows as an expression of inter-industry relatedness. These papers assume that human capital is to some extent industry specific and that, therefore, that most labor flows will take place between industries with similar human-capital requirements. This has resulted in labor-flow-based skill-
relatedness measures (Neffke and Henning, 2013), which have been used in a variety of papers (e.g., Timmermans and Boschma, 2012; Boschma et al., 2014; Diodato and Weterings, 2015). In accordance with this literature, a second set of stylized facts analyzes inter-industry labor flows through the prism of skill relatedness. In particular, we are interested in four issues. First, how stable are skill-relatedness structures? Although inter-industry relatedness patterns may change in the long-run, a large year-on-year volatility in skill relatedness would be undesirable. Second, how general are skill-relatedness measures? That is, do different types of workers exhibit different skill-relatedness patterns? Third, given that many workers tend to search for jobs in their own region, to what extent do skill-relatedness measures just reflect industrial colocation patterns? And, fourth, what is the predictive validity of skill-relatedness measures vis-à-vis alternative relatedness measures?

We derive the stylized facts from Germany’s social security records between 1999 and 2008, which cover over 80% of the working population. We find that, although workers often do switch industries, even at a very high level of aggregation (stylized fact 1), labor flows are highly structured (stylized fact 2). In particular, 50% of the workers leaving an industry are subsequently employed in related industries that together represent less than 4% of total German employment. Moreover, the underlying network of labor flows is largely independent of a worker’s occupation: job switches of workers in different occupations tend to concentrate in the same industry pairs. This suggests that, independently of any occupational specificities, job switches are guided by a non-negligible industry-specific component in human capital. When we turn to labor flows as a measure of inter-industry relatedness, we find that the

¹ Inter-industry labor flows express to what extent different industries share the same workers. Although this may reflect social networks and workers’ preferences, Neffke and Henning argue – among other things, quoting evidence from skill surveys – that the main determinant of the underlying structure of inter-industry labor flows is skills. This is why they term the relatedness index in their paper skill relatedness. Like most subsequent papers, we follow this convention and refer the reader for a full discussion to Neffke and Henning (2013).
derived skill-relatedness index is remarkably stable (stylized fact 5), general across occupations and wage levels and similar in former East and West Germany (stylized fact 4). Furthermore, given that intra-regional flows follow a similar skill-relatedness structure as long-distance flows (stylized fact 3), skill-relatedness is not simply a reflection of the industrial composition of local economies. Moreover, in a direct comparison, our labor-flow-based measure outperforms commonly used alternative relatedness measures in predicting entry and growth rates of local industries (stylized fact 6). Finally, skill-related industries have uncorrelated growth patterns, suggesting that skill-relatedness should typically not impede the reallocation of labor from shrinking to growing industries (stylized fact 7). The skill-relatedness matrices are collected in an online appendix. Although the paper only describes skill-relatedness for the industrial classification used in Germany in the period 1999 to 2008, to facilitate future research, the online appendix contains matrices for various industrial and occupational classification systems that have been in use between 1975 and 2014.²

The paper is structured as follows. In section 2, we discuss the literature on human capital specificities and job switches and the literature on inter-industry relatedness measures. Section 3 describes the data. In section 4, we develop a number of statistical tools to analyze labor flow networks and present the stylized facts uncovered through these tools. Section 5 discusses future research and concludes.

2. Literature review

Human capital specificity and job switching

---

² A link to these matrices, as well as detailed description of the procedure used to create these matrices, is provided on the first author’s personal website.
Human capital and skills are pivotal inputs in today’s production processes, which is why a firm’s workforce is regarded as an exceedingly important competitive asset (Porter, 1987; Grant, 1996; Grant and Spender, 1996). Moreover, today’s workforces are highly specialized: individual workers often invest heavily in education and training to acquire specific skills that allow them to carry out specific tasks, running the gamut from engineering to financial management and from construction work to food preparation. Because workers specialize, their human capital is often held to be specific to the firm where they work (Becker, 1964), to an industry (Neal, 1995; Parent, 2000; Sullivan, 2010) and to occupations and tasks performed (Poletaev and Robinson, 2008; Gathmann and Schoenberg, 2010). However, there is considerable debate about which of these dimensions of skill specificity dominate. For instance, Kambourov and Manovskii (2009) study the value of occupation tenure and industry tenure and report that only the former is rewarded with higher wages, casting doubt on the existence of industry specificities in human capital. In contrast, Sullivan (2010) shows that industry tenure sometimes does reap high rewards that are not explained by occupation tenure, but that these rewards depend on the occupation.

We propose to approach the question of whether human capital has an industry-specific component differently. The starting point is that job-switching patterns contain valuable information on the nature of human capital. After all, job switches will render some human capital redundant, whenever the old and the new job require different skills. To avoid such human capital depreciation, workers will predominantly switch to jobs that allow them to reuse their skills. This suggests that overlap in industries’ human capital requirements, or, more accurately, the absence of such overlap, should constrain inter-industry labor movements. Therefore, inter-industry job switches contain information on which industries value similar skills and know-how, indirectly shedding light on the existence of human capital specificities.

Job switches have been studied in great detail in labor economics. Much of this literature’s interest in labor flows is driven by their role in readjusting the allocation of labor across firms by matching
workers to jobs. For instance, in an overview paper, Davis and Haltiwanger (1999, p. 2713) motivate the widespread interest in labor flows as follows: “[T]he extent to which the reallocation and matching process operates smoothly determines, in large measure, the difference between successful and unsuccessful economic performance.” Research in this tradition has resulted in numerous papers summarized in various reviews (Davis and Haltiwanger, 1999; Farber, 1999; Davis et al., 2006). These papers mainly study the rate at which jobs are created and destroyed, the rate at which workers change jobs, and whether workers do so voluntarily or involuntarily. Other important questions are how job-switching rates develop over the business cycle and what this means for unemployment dynamics and labor-market institutions.

Although most of this work in labor economics has focused on the question of how often and why workers change jobs, more recently, scholars have turned to the issue of workers switching industries. These studies find that, in the United States, workers change 1-digit industries at relatively high rates of between 12% (Kambourov and Manovskii, 2008) and 20% (Parrado et al., 2007) a year. These findings may be interpreted as a sign that workers are not strongly constrained in their movements across industries, implying that human capital is not highly specific to an industry. But this conclusion rests on the assumption that the hierarchy of the industry classification system groups industries by their human capital requirements, whereas a similar notion - that industry classification systems reflect similarities in use of strategic resources - has been heavily criticized in strategic management (Robins and Wiersema, 1995; Bryce and Winter, 2009; Neffke and Henning, 2013). In the empirical section, we will show that such assumptions are indeed not just problematic, but that evidence based on them is misleading.

Knowledge spillovers

Labor mobility has also been studied in a different context, namely, as a mechanism for knowledge diffusion. Accordingly, workers who switch jobs do not just reallocate labor, but often also carry with them
valuable knowledge, expertise and networks from their previous employer (Almeida and Kogut, 1999; Cantner and Graf, 2006; Storz et al., 2015). Moreover, because most individuals change jobs within their regions, such knowledge spillovers tend to be spatially constrained (e.g., Agrawal et al., 2006). However, the scope for knowledge spillovers may also be limited by human capital specificities. In particular, economic activities that employ radically different skills and knowhow will not only have limited scope for labor mobility among them, it is also less likely that knowledge and technologies can be fruitfully shared. As highlighted in the literature on the optimal cognitive distance among economic activities (Nooteboom et al., 2007), many industries’ knowledge bases are so different that there will be little scope for knowledge spillovers.

**Related diversification**

The considerations on the existence of industry-specificities in human capital and of an optimal cognitive distance both suggest that inter-industry labor flows will mostly occur among related industries. Conversely, this would imply that inter-industry labor flows might be used to measure inter-industry relatedness. The concept of inter-industry relatedness has played an important role in the literature on firm diversification (Penrose, 1959, Teece et al, 1994) and, more recently, also in economic geography and the literature on geographical clusters of firms. For instance, Porter (1998) identifies local clusters of related economic activities as important sources of competitive advantage, with Silicon Valley as the archetypal example. Although originally stressing local value chains, the cluster concept has evolved to include more general linkages that “create externalities of various types” (Porter, 2003). To measure such linkages, scholars have used information on the co-location patterns of industries (Porter, 2003) and the co-occurrence of products in countries’ export portfolios (Hidalgo et al., 2007). Although these and other relatedness measures have proven highly predictive of the growth of local industries (Boschma et al., 2012; Delgado et al., 2010; 2014; Essletzbichler, 2013; Hausmann et al., 2014; Neffke et al., 2011; Rigby,
recently, labor-flow based inter-industry labor flows have gained much support. For instance, Greenstone et al. (2010) show that large-plant openings create spillovers to local firms, but in particular to firms in industries that are related to the new plant’s industry as measured by labor flows. Similarly, labor-flow-based inter-industry relatedness measures have been used in other studies in economic geography (Dauth, 2010; Timmermans and Boschma, 2014), trade (Kaplan et al., 2011), strategic management (Neffke and Henning, 2013) and entrepreneurship research (Costa and Baptista, 2011). To our knowledge, however, there has hitherto not been any detailed investigation of the structure of inter-industry labor flows that would assess the legitimacy of such labor-flow based relatedness measures.

3. Data

Our data are constructed from Germany’s social security records as compiled in the Historic Employment and Establishment Statistics (HES) database. The HES offers a large set of demographic and employment characteristics, such as a worker’s daily wage, occupation, work status (i.e., apprentice, part-time worker, full-time worker), gender and age. Furthermore, the industry and location of each individual’s work establishment are known.

We limit the analysis to full-time employees aged 18 to 65. Furthermore, we exclude apprentices and volunteers because they are still investing in education to acquire skills. Because upper limits to social security contributions result in right-censored wage information, we impute wages whenever they exceed the contribution limits following Gartner (2005). Due to changes in the industry-classification system (see

---

3 Bender et al. (2000) provide a detailed description of this database.

4 We deflate wages to 2005 EUR.

5 We drop workers employed through employment agencies, because we don’t have information on the actual industry or region in which these individuals work.
for a more detailed description Appendix A), we confine our analyses to the years 1999 to 2008. This results in a final data set on, on average, about 20 million workers a year.

**Definition of labor flows**

We use the HES to construct inter-industry labor flows. Labor flows arise when workers switch establishments from one year to the next. Workers who enter or exit the social-security data in this period are ignored in these flows. As a consequence, these labor flows predominantly reflect job-to-job switches, although some of them may feature a short (unobserved) unemployment spell in between the two observed jobs. Although this eliminates some involuntary job switches that lead to prolonged unemployment spells, we have no reliable way of differentiating between voluntary and involuntary job switches. Moreover, establishment identifiers in the HES are not perfectly reliable. For instance, spin-offs, mergers, break-ups or mere recodings all would introduce new identifiers without signaling de-novo entries. Hethey and Schmieder (2010) find that for only 35% to 40% of all establishments with over three employees a new (or a disappearing) establishment identifier can be interpreted unambiguously as an entry (or as an exit). In the other cases, workers move in larger blocks from one establishment to another. To avoid that such spurious identifier changes contaminate our labor-flow measurements, we remove 531,000 job switches (27.5%) from a total of the 1.8 million yearly job switches.\(^6\)

**Labor-flow segments**

In the empirical section of this paper, we decompose flows into different segments. We introduce three kinds of segmentations, one based on the geography of flows, one that captures workers’ skills and one that distinguish between an eastern and a western German labor market. Note that the latter two

\(^6\) See Appendix B for a more detailed description of the identification and elimination of spurious job switches.
segmentations, which use job characteristics, are based on the job that represents the origin of a job switch, not the destination.

The first segmentation is motivated by the potential concern that the structure of inter-industry labor flows is governed by the availability of local jobs. In that case, inter-industry labor flows could simply be an expression of the co-location patterns of industries. We look into this by comparing job switches over short distances to long-distance switches, i.e., switches for which the old and the new job are at least 100 kilometers apart.\footnote{Distances between old and new jobs represent road distances between the centroids of the districts (Kreise) in which the corresponding establishments are registered. Given that only five percent of German employees commute over distances greater than 50 km (Winkelmann, 2010), we postulate that distances of over 100 km typically require a worker to relocate.}

Second, to proxy workers’ human capital levels, we segment the labor market into workers who earn below and those who earn above the median wage\footnote{We use a worker’s wage to proxy worker quality, instead of for instance, his or her education and experience or, its complement, a Mincer residual, because wages will reflect both observed and unobserved quality characteristics.} in their industries. To explore whether the type of human capital matters, we next separately analyze workers in eight broad occupational groups that are associated with different broad sets of tasks. In particular, we distinguish among managers, sales-related employees, accountants, information technology (IT) workers, office clerks, cleaners, security personnel and other workers (see Appendix C). These occupations were chosen because of their relative ubiquity across a variety of industries.\footnote{This ensures that we observe flows that are not limited to a small subset of sectors. For instance, if we were to use occupations such as glass makers, we would not expect any labor flows to originate from banks, given that such occupations will simply not exist in that industry.}

Third, we investigate whether there are regional differences in labor-flow patterns by splitting the German labor market along the former border between East and West Germany.\footnote{To ensure that our results are not driven by the extensive outmigration from the East after Germany’s reunification, we exclude flows between East and West Germany in these analyses.}
4. Results

Below, we derive a number of stylized facts using the data described in the previous section. We first describe the general structure of the labor flows in terms of the amount of job switches and the degree to which labor flows concentrate in relatively few industry pairs to shed light on the specificity of human capital. Next, we analyze the general structure underlying these flows by plotting the skill-relatedness network and comparing the skill-relatedness matrices for different labor-market segments. Finally, we turn to the question of how skill relatedness affects local labor markets by estimating local industry-growth regressions and determining the extent to which the limited mobility of workers across industries could hinder an efficient reallocation of workers from shrinking to growing industries. These analyses draw on a number of statistical tools and indicators for which the intuition is described in the main text, Details and derivations are provided in the appendix to this paper.

Cross-industry labor-flow patterns

At their coarsest level of aggregation, industries in the European NACE Revision 1.1 classification are divided into sections. Sections consist of several sub-sections, which themselves are made up of 2-digit industries. These 2-digit industries are further subdivided into 3-, 4-, and 5-digit industries. Table 1 summarizes average yearly labor flows across industries at these different levels of aggregation. The first column of Table 1 reports figures for Germany as a whole. Of all workers who change jobs, 73.4% change industries at the 5-digit level, the most disaggregated level available. In line with Kambourov and Manovskii’s (2008) finding that much industry switching takes place across highly aggregated industries,
we find that 58.7% of these 5-digit industry switchers switch industries at the section level, the most aggregated industry grouping in the NACE 1.1 classification.\textsuperscript{11}

Distinguishing flows by their labor-market segment of origin, we find that workers in the high-income segment switch industries less often and undertake less drastic switches than low-wage workers do. For instance, 8.1% of high-wage workers who switch industries stay in their 4-digit industries (column 2), against only 6.2% for low-wage workers (column 3). Workers in eastern Germany switch industries somewhat less than their colleagues in the west (columns 4 and 5). Columns 6 and 7 show that jobs switches over distances below 100 km (“local” labor flows) display patterns that are very similar to those that involve distances of over 100 km (“long-distance” labor flows). Differences by occupation are larger.\textsuperscript{12} Workers in lower-skilled occupations such as cleaning (78.8%) and security (73.9%) cross section borders much more often than workers in higher-skilled occupations, such as management (58.4%), accountancy (58.4%) and IT (58.5%), who tend to switch to industries that are classified more closely to their old jobs. This suggests that moving to a radically different industry is less attractive when human capital requirements are higher. However, in spite of these differences among labor-market segments, overall, the following stylized fact emerges from Table 1:

\textsuperscript{11} Compared to a random benchmark in which workers choose new industries with a probability equal to the industry’s employment share in the overall economy, workers tend to remain in their industry relatively often (in the overall economy, the likelihood of staying in one’s industry is 39 times higher than this random benchmark). Moreover, switches to industries that are nearer in the classification system typically exceed this benchmark by more. For instance, in Germany as a whole, switches within 4-digit industries exceed the random benchmark by a factor 15, whereas switches to other sections happen at rates of 0.7 times the random benchmark. An exception are cleaners, whose switching rates exceed the random benchmark when they change sections, but not when they stay within sections, subsections or detailed industries.

\textsuperscript{12} Workers in security and cleaning jobs seem to switch industries less often than other workers. However, this is misleading because cleaning and security employees work predominantly in the cleaning and the security industry. As a consequence, most job switches occur within these two 5-digit industries. Once cleaners and security guards cross industry boundaries, their labor flows look much less structured than the ones in other occupations.
Stylized fact 1: Workers frequently cross industry boundaries, even at the most aggregate level of the industry classification system. Moreover, such “distant” moves are most frequent for workers with low wages and in low-skill occupations.

TABLE 1 (FLOWS) ABOUT HERE

Flexibility

The fact that workers tend to cross boundaries between highly aggregated sectors does not necessarily mean that labor flows are unstructured. Indeed, 3.3% of all possible pairs of 5-digit industries account for 80% of all job switches and in 56% of pairs we do not observe any job switches whatsoever in the nine-year period we study. How random then are labor flows? To answer this question, we calculate for each industry the minimal set of industries that together absorb 50% of the industry’s labor outflow. Let $F_{ij}^s$ be the labor inflow into industry $j$ that originates from industry $i$ in segment $s$. Furthermore, let $\alpha_{ij}^s$ be:

$$
\alpha_{ij}^s = \frac{F_{ij}^s}{N_j^s}
$$

where $N_j^s$ is the number of workers in industry $j$ in segment $s$. $\alpha_{ij}^s$ can be thought of as the per-employee-rate at which industry $j$ absorbs workers from industry $i$. Next, for each industry of origin, we sort destination industries in descending order of this ratio: $\alpha_{i1}^s > \alpha_{i2}^s > \cdots > \alpha_{in}^s$, where $n$ is the number of industries in the economy. Throughout this paper, we will use a dot (".") to denote summation over an omitted dimension. Hence, $F_i^s = \sum_j F_{ij}^s$ represents the total labor outflow in a segment $s$ from industry $i$. Furthermore $N_i^s = \sum_j N_j^s$ represents the total employment in this segment. We define $FLEX_i^s(q)$ as:

$$
FLEX_i^s(q) = \sum_{j=1}^{k} \frac{N_j^s}{N_i^s} \quad \text{with} \quad k = \arg\min_{k'} \left( q < \sum_{j=1}^{k'} \frac{F_{ij}^s}{F_i^s} \right)
$$
In other words, $FLEX_i^s(q)$ gives the minimum employment share of a set of 5-digit industries that together absorb at least $q\%$ of all workers who leave industry $i$ in segment $s$.\(^{13}\) To determine the flexibility of workers in an entire labor-market segment, we calculate the outflow-weighted average of $FLEX_i^s(q)$ across all industries of origin:

$$FLEX^s(q) = \sum_i \frac{F_i^s}{F^s} FLEX_i^s(q)$$

where $F^s = \sum_i F_i^s$ represents the total inter-industry labor flows that originate in labor-market segment $s$.

$FLEX^s(q)$ can be calculated both with and without within-industry labor flows, that is with or without workers who change establishments, but not industries. Figure 1 depicts the average $FLEX^G(q)$ over the period 1999 to 2007 against $q$, where $G$ represents the “segment” that contains the entire German labor market. $FLEX^G(q)$ is highly convex, meaning that the bulk of labor outflows are absorbed by a small part of the economy. Because some of the curvature of $FLEX^G(q)$ would also have occurred, had flows simply followed employment, we create a benchmark based on simulated flows. In this simulated benchmark, switchers in industry $i$ randomly choose a destination industry $j$ with a probability equal to $j$’s employment share in the overall economy: $p_j^s = \frac{N_j^s}{N^s}$. The resulting null-model curve is depicted as a dotted line.

FIGURE 1 (FLEX) ABOUT HERE

---

\(^{13}\) Where necessary, we interpolate $\sum_{i=1}^{k} \frac{N_i}{N}$ between $k$ and $k + 1$. 15
Table 2 compares the flexibility of workers in different labor-market segments by providing $FLEX(50)$ values for each segment excluding (column 1) and including (column 2) within-5-digit-industry flows. The number in parentheses shows the corresponding random benchmark.

**TABLE 2 (FLEX) about here**

In general, labor flows are tightly structured, especially if we include within-industry flows.\(^\text{14}\) For Germany as a whole, 50% of all workers move to sets of industries\(^\text{15}\) that represent only 3.4% (5.9% if within-industry flows are excluded) of the economy. High-wage workers, with a $FLEX(50)$ of 2.6% (4.2%), are much less flexible than low-wage workers, with a $FLEX(50)$ of 3.4% (5.9%). Given that low- and high-wage groups have virtually identical null-model predictions, their estimates can be readily compared. For other labor-market segments, null-model predictions vary widely, which complicates comparisons across groups. However, given that, in all segments, $FLEX(50)$ values are well below their null-model predictions, we arrive at the following stylized fact:

**Stylized fact 2:** Labor flows are channeled along tight paths: most workers are absorbed by sets of industries that represent only a small fraction of total employment in the economy.

Stylized fact 2 is a first indication that human capital is highly specific to a worker’s job. Indeed, most industry switches are all but completely rejected by workers. In the next section, we study to what extent this pickiness of workers when it comes to choosing a new industry can be interpreted as an expression of industry-specific human capital.

\(^{14}\) The exception is occupational segments that are dominated by a single industry (as in cleaning and security jobs), where most of the flows take place among firms within that 5-digit industry. Here including the within-industry flows substantially decreases the estimated flexibility.

\(^{15}\) Note that these sets differ by industry, that is, each industry will have a different set of industries that absorb most of its worker outflow.
The skill-relatedness structure of labor-flow matrices

So far, we have documented patterns in raw labor flows. However, the size of labor flows will depend on the sizes and flow rates (i.e., the fraction of employees switching jobs) of the industries involved. To isolate the structure underlying inter-industry labor flows, we calculate the ratio between the observed volume of labor flows, and the one that would be expected from industries’ flow rates. If workers switched industries with probabilities proportional to the total outflow of the industry of origin, $F_i^S$, and the total inflow into the destination industry, $F_j^S$, the expected labor flow between $i$ and $j$ is given by $F_{ij}^S = \frac{F_i^S F_j^S}{F_j^S}$ and the ratio of observed to expected flows by:

$$R_{ij}^S = \frac{F_{ij}^S}{F_{ij}^S}$$

Equation (1) yields $R_{ij}^S$ values that have slightly higher year-on-year correlations, suggesting a somewhat higher consistency. In practice, however, both baselines give virtually the same results, with rank correlations of around 0.98. An intermediate solution is provided in Neffke and Henning (2013), who use regression analysis to predict expected flows from general industry level characteristics. Also this method yields very similar results.
which maps $R^S_{ij}$ values between 0 and 1 onto the interval $[-1,0]$ and values from 1 to infinity onto the interval $[0,1)$. As a result, $R^S_{ij}$ is symmetric around zero. \footnote{In particular, observed flows that exceed expected flows by a factor $a$ translate into $R^S_{ij} = \frac{a-1}{a+1}$, whereas the opposite – expected flows exceeding observed flows by a factor $a$ – yields $R_{ij} = \frac{1}{a} - \frac{1}{1+a} = \frac{a-1}{a+1}$.} Although the flow patterns may reflect other factors than just an overlap in industries’ skill requirements, following Neffke and Henning (2013), we interpret these inter-industry labor flows as an expression of skill similarities and refer to $R^S_{ij}$ as the \textit{skill relatedness} of $i$ to $j$ in segment $s$ and call $i$ and $j$ \textit{skill related} if $R^S_{ij} > 0$.

Figure 2a and 2b show two visualizations of the $R$-matrix for the German labor market as a whole. Figure 2a shows a heat map for skill-relatedness estimates among all possible 5-digit industry pairs, with rows and columns sorted by an average-linkages hierarchical clustering algorithm. The dark squares along the figure’s diagonal indicate that the matrix exhibits a fair degree of clustering, i.e., certain sets of industries are densely connected among each other. However, there are also links across these clusters, as evidenced by various dark off-diagonal areas.

To get an impression of which industries are connected, Figure 2b shows the network spanned by just the top 651\footnote{We display three times as many links as nodes, which, as a rule of thumb, yields networks that are not overly cluttered.} values in the $R$-matrix, instead of depicting the entire skill-relatedness matrix. In this figure, nodes represent 3-digit\footnote{Aggregating industries at the 3-digit level allows us to label the majority of nodes.} industries (colored according to the NACE sections to which they belong). The size of a node represents the corresponding industry’s average employment in the period we study. The layout of the network is based on an algorithm that aims at grouping closely related industries together such that nodes that cluster in the graph generally correspond to sets of skill-related industries.

FIGURE 2b (IND SPACE) about here
Industries tend to cluster by section. For instance, we find a cluster of dark-blue textiles and leather industries (center-left), a cluster of hotels and transport-related services (green, upper-right), and a large metals-and-electronics cluster (blue, top-left). At the same time, there are large labor flows among industries with different colors, i.e., of different sections. However, these links across sections often connect industries that are intuitively related. For instance, high-technology manufacturing and service industries are connected in a cluster that links the computer and telecommunications equipment industries of the blue manufacturing section with the software consulting and data processing industries of the orange business-services section. Similarly, we find strong links between the manufacturing industries of printing and publishing and creative services like radio & TV and advertising.

**Comparing skill relatedness across labor-market segments**

Does the network in Figure 2 depict a general structure or does skill relatedness differ by labor-market segment? To answer this question, we need to compare different skill-relatedness matrices to one another. We do this by first stacking all columns of a skill-relatedness matrix into one long vector. Next, we calculate the correlations among such vectors for different segments.

The estimated correlations are surprisingly low, typically between 0.3 and 0.5. However, even if we compare the skill-relatedness matrix for the same segment in two consecutive years, correlations barely exceed the 0.5 mark. Although, in the long run, skill relatedness may change as technologies shift, it is implausible that it changes much on such a short horizon. This suggests that skill-relatedness matrices are estimated with a substantial amount of noise. Indeed, if we assume that skill relatedness does not change from one year to the next, a year-on-year correlation of 0.5 suggests that 75% \((1 - 0.5^2)\) of the variance in the estimated skill relatedness can be attributed to random noise. As a consequence, correlations involving skill relatedness will have a strong attenuation bias, i.e., they will be biased towards zero.
To resolve this, let skill-relatedness estimates be composed of two components: the (unobserved) actual skill relatedness and measurement error. Formally, we write the stacked vector of skill-relatedness estimates for segment $s$ in year $t$, $\hat{r}_{st}$, as the sum of a time-invariant, real skill-relatedness vector, $r_s$, and a year-specific measurement error component, $\epsilon_{st}$:

$$\hat{r}_{st} = r_s + \epsilon_{st}$$  \hspace{1cm} (1)

If we assume that $\epsilon_{st}$ is distributed identically and independently, equation (2) implies that the bias in the correlation between two observed skill-relatedness vectors, $\hat{r}_{st}$ and $\hat{r}_{st}'$, can be reduced by averaging the skill-relatedness estimates of different years. However, because we only have nine yearly estimates, averaging $\hat{r}_{st}$ will not eliminate the attenuation bias completely. Correlations between time-averaged skill-relatedness vectors, therefore, represent lower bounds of correlations of vectors of actual skill relatedness.

An alternative approach is to tackle the measurement error directly. This is possible, because our yearly estimates of skill relatedness, in principle, represent nine independent estimates of skill relatedness. The reason is that we observe each worker only once a year. Therefore, a worker who moves out of an industry in year $t$ cannot undertake that same move again in year $t + 1$. As a consequence, for each industry pair $(i, j)$, flows in consecutive years are necessarily composed of disjoint worker sets.\(^{20}\)

Consequently, we can use the bias-correction method introduced by Spearman (1904):

---

\(^{20}\) The assumption that yearly relatedness estimates are uncorrelated is still quite strong. For instance, workers may imitate coworkers’ previous moves and shocks that make an industry switch attractive in one year may show some degree of autocorrelation. However, such considerations suggest that year-on-year correlations of the error term should be positive. As a consequence, the denominator in bias-correction equation (3) would be too large, meaning that, in the worst case, our estimates would underestimate true cross-segment correlations.
\[ Corr[r_s, r_{s'}] = \frac{Corr[r_{s,t}, r_{s',t}]}{\sqrt{Corr[r_{s,t}, r_{s,t+1}]} \cdot \sqrt{Corr[r_{s',t}, r_{s',t+1}]]} \]  

That is, the true correlation between skill-relatedness vectors for segments \( s \) and \( s' \) can be estimated by dividing the correlation between observed skill-relatedness vectors by the square root of the correlations between two consecutive skill-relatedness measurements in each segment. Detailed derivations are provided in Appendix D.\textsuperscript{21}

In the analyses below, we aggregate industries to the 3-digit level to strike a balance between the need to use relatively homogeneous industries and to maintain a large enough ratio of labor flows to industry pairs to calculate sufficiently precise skill-relatedness estimates. Tables 3 to 5 summarize, for each of the three segmentations described in section 3, the correlations among skill-relatedness matrices of different labor market segments. The first row in these tables reports the correlation for a single segment’s skill relatedness in two consecutive years.\textsuperscript{22} The higher this year-on-year own-correlation, the less noisy estimates are. The remaining rows describe correlations between the skill-relatedness measurements in a given pair of segments.\textsuperscript{23} To calculate the bottom value in these rows, we first average skill relatedness for the segments across years and then calculate the correlation of these averages. This provides a lower bound on the true correlation of two relatedness types. The upper value reports the average bias-corrected correlations using equation (3).

One potential concern is that skill-relatedness does not measure similarities in skill requirements, but simply reflects industries’ co-location patterns. If this were the case, we would expect that labor flows

\textsuperscript{21} The appendix also provides outcomes using a third bias-correction method. This approach combines the two methods described above and yields very similar results as the bias-correction method of equation (3).

\textsuperscript{22} We first calculate correlations for all eight pairs of consecutive years and then take the average.

\textsuperscript{23} Because we can only estimate the relatedness among industries with at least some inflow or outflow in each of the two labor-market segments, the number of observations varies across the tables’ cells.
taking place within a region would be structurally different from those that cross into other regions. However, Table 3 shows that, at a bias-corrected correlation of 0.87, local and long-distance labor flows exhibit very similar skill-relatedness estimates. This suggests that the structure of the skill-relatedness network is not driven by industry co-location patterns.

**TABLE 3 (CORR REG) about here**

**Stylized fact 3:** Given that local and long-distance flows exhibit very similar skill-relatedness structures, skill-relatedness estimates are not mainly driven by industry co-location patterns.

Table 4 shows that, in eastern as well as in western Germany, workers with different wage levels have almost identical skill-relatedness matrices: all bias-corrected correlations are well above 0.9 and even without bias correction, we observe correlations of at least 0.8. Although differences between eastern and western Germany are slightly larger, with bias-corrected correlation estimates typically above 0.75, also these differences are relatively small.

**TABLE 4 (CORR WAGES) about here**

Table 5 reports correlations of skill relatedness in different occupational segments. With the exception of cleaning and security personnel, for whom bias-corrected correlations with the other occupational groupings hardly ever exceed the 0.5 mark, the different occupational groupings exhibit very similar skill-relatedness matrices.\(^{24}\) Management occupations display skill-relatedness matrices that are almost identical to those of sales people (bias-corrected estimate: 0.92), accountants (0.86), office clerks (0.91)

\(^{24}\) Apparently, cleaners and security guards' job transitions are quite distinct from those of the other groups. Interestingly however, with a bias-corrected correlation of 0.79, cleaners and security guards display very similar skill-relatedness matrices to one another. A closer inspection of their relatedness matrices suggests that even in these jobs, where industry-specificity of skills is arguably low, workers do not switch industries randomly. For instance, in both groups, we find strong connections among various construction industries.
and, to a somewhat lesser extent, IT specialists (0.83). Indeed, even without correcting for measurement error, all correlations in the first five occupational segments exceed 0.6. This stability across occupations is remarkable. Even though managers and IT specialists will carry out different tasks, they tend to switch jobs among the same industries. In sharp contrast to the findings of Kambourov and Manovskii (2009) and Poletaev and Robinson (2008) that industry specificity is scarcely important after taking occupational specificity into account, the fact that inter-industry flow patterns are similar for different occupations suggests that human capital does have an industry-specific component that is quite independent of a worker’s occupation.

**TABLE 5 (CORR OCC) about here**

Taking together the results in Tables 4 and 5, we arrive at the following stylized fact:

**Stylized fact 4:** Workers with different levels and types of skills have similar skill-relatedness networks, i.e., the same industries are connected by labor flows, regardless of the skills of workers involved.

Using the bias-correction method of equation (3), we can also assess the extent to which skill relatedness changes over time. The bias-corrected correlation between skill relatedness in the first and last year of our data set is 0.9 or higher in all segments (Table 6). This shows that, if at all, skill relatedness changes very slowly over time.

**TABLE 6 (CORR TIME) about here**

**Stylized fact 5:** There is no indication of rapid change in skill relatedness.

---

25 The measurement-error correction is in this case based on equation (4) and uses the year-on-year same-segment correlations between 1999/00 and 2000/01 and between 2006/07 and 2007/08 in the denominator.
Skill relatedness and the growth of local industries

As explained in section 2, various measures of inter-industry relatedness have been found to predict the diversification paths of regions. In this subsection, we test the predictive validity of skill-relatedness within this context, by investigating how skill-relatedness compares to alternative relatedness measures as a predictor of local industry growth. In particular, we assess to what extent the presence of related industries predicts the growth and entry of local industries in German planning regions (Raumordnungsregionen). We define related industries on the basis of three different relatedness measures: skill relatedness, a co-location-based measure as in Porter (2003) and Hidalgo et al. (2007), and an input-output-based measure. All three measures are created using data for the period before 2003. In contrast, the analyzed growth patterns refer to the period 2003-2008.

As a measure of skill relatedness, we take the average skill-relatedness for Germany as a whole in 1999/00, 2000/01 and 2001/02. Next, we symmetrize the resulting matrix as follows:

$$SR_{ij} = \frac{R_{ij} + R_{ji}}{2}$$

Co-location-based relatedness is calculated as the correlation between the regional employment vectors of two industries:

$$CL_{ij} = \frac{1 + \text{corr}(E_{ir}, E_{jr})}{2}$$

(3)

where $E_{ir}$ represents the employment in industry $i$ and region $r$ in the year 2002. The transformation in (4) maps the correlation onto the interval [0,1]. $CL_{ij}$ increases as the geographical distribution of employment in industries $i$ and $j$ becomes more similar.
For input-output relatedness, we use the German 2-digit input-output matrix of the year 2003 as provided by EUROSTAT.\(^{26}\) The input-output matrix records for each pair of industries \((i, j)\) the value of industry \(i\)'s outputs purchased by industry \(j\), \(V_{ij}\). To establish the strength of input-output linkages between two industries, we express \(V_{ij}\) once as a percentage of all intermediate inputs into industry \(j\) and once as a percentage of all intermediate outputs of industry \(i\). We then repeat this for the reverse flow, the value of intermediates sold by industry \(j\) to industry \(i\). Input-output relatedness is now defined as the average of these four figures:

\[
IO_{ij} = \frac{1}{4} \left( \frac{V_{ij}}{V_{i}} + \frac{V_{ji}}{V_{j}} + \frac{V_{ji}}{V_{i}} + \frac{V_{ij}}{V_{j}} \right)
\]

where a dot (\(\cdot\)) once more denotes summation over the omitted dimension. For all relatedness measures, we define industries to be unrelated to themselves: \(SR_{ii} = CL_{ii} = IO_{ii} \equiv 0\).

We use these relatedness measures to quantify how well a local industry’s related industries are represented in the region. To do so, we calculate for each industry \(i\) the weighted average employment of all other industries, \(j \neq i\), in the region, using the relatedness between \(i\) and \(j\) as weights:

\[
E_{iRt}^{REL} = \sum_{j \neq i} \frac{REL_{ij}}{\sum_{m \neq i} REL_{im}} E_{jRt}
\]

where \(E_{jRt}\) is industry \(j\)'s employment in region \(r\) and year \(t\) and \(REL\) is either \(SR\), \(CL\) or \(IO\). Next, we use these variables to predict annualized growth rates of existing local industries and the entry of new industries in a region. These regressions are limited to local industries in the traded, non-resource-

\(^{26}\) To be precise, we use the domestic section of the product-based symmetric input-output table, ignoring international trade and sales to end-consumers.
based, private sector. Moreover, for the growth regressions, we only use industries that exist at the start of the period, estimating (cross-sectional) regression equations of the following type:

\[
\log \left( \frac{E_{it}}{E_{it+\tau}} \right)^{\frac{1}{\tau}} = \gamma \log(E_{it}) + \beta_{SR} \log(E_{it}^{SR}) + \beta_{CL} \log(E_{it}^{CL}) + \beta_{I0} \log(E_{it}^{I0}) + \eta_i + \rho_r + \epsilon_{it}
\]  

(4)

In equation (5), \(\gamma\) estimates potential mean-reversion effects. \(\eta_i\) and \(\rho_r\) represent industry- and region-specific effects respectively. Furthermore, \(t\) is the base year of the growth rate and \(\tau\) represents the time horizon over which growth is measured such that the dependent variable reflects annualized growth.

To analyze the entry of new local industries, we run Linear Probability Models (LPMs). That is, our dependent variable is a dummy variable \(ENTRY_{it,\tau}\) that has a value of 1 if an industry \(i\) that did not yet exist in year \(t\) enters region \(r\) within the next \(\tau\) years:

\[
ENTRY_{it,\tau} = \beta_{SR} \log(E_{it}^{SR}) + \beta_{CL} \log(E_{it}^{CL}) + \beta_{I0} \log(E_{it}^{I0}) + \eta_i + \rho_r + \epsilon_{it}
\]

(5)

Like equation (5), equation (6) describes a cross-sectional analysis with industry and region fixed effects. However, because we only use local industries that do not yet exist in year \(t\), there is no mean-reversion term in this equation.

Tables 7a and 7b report estimates for (5) and (6) with base years \(t = 2003\) and growth and entry defined over a one-year and over a five-year period. Table 7a aggregates industries to the 3-digit level, whereas Table 7b presents estimates for 4-digit industries. In all but one model, only the coefficients associated with skill-related employment are statistically significant. The estimated coefficients suggest that a 10%

---

27 That is, we exclude industries in the 2-digit NACE classes 01-14 (agriculture, fishing and mining), 40-59 (non-traded services), and 75-99 (public sector and miscellaneous industries).

28 The use of LPMs allows controlling for the same full set of region and industry specific effects as used in the growth regressions. In contrast, nonlinear models such as probit or logit models cannot accommodate such a large number of regressors.

29 Results are qualitatively similar for other base years and time windows (results available on request).
rise in skill-weighted average employment is associated with a between 1.5% and 2% increase in growth rate and between 0.5 \((0.051 \times \ln(1.1))\) and 1.0 \((0.104 \times \ln(1.1))\) percentage points higher entry rates. This shows that industries tend to enter and grow faster in regions with large amounts of skill-related employment. In contrast, controlling for skill-related employment, employment that is related according to co-location or input linkages typically does not show any association with growth and entry rates.

**TABLE 7a AND 7b ABOUT HERE**

This is not to say that co-location and input-output relatedness indicators have no merit. For one, these analyses are complicated by collinearities among the various \(\log(E_{iirt}^{REL})\) terms.\(^{30}\) Indeed, if we regress growth and entry rates on each indicator separately all indicators exhibit some explanatory power. For another, we find stronger results for co-location-based relatedness (but not input-output based relatedness) if, instead of using the relatedness-weighted average employment, we measure related employment as all employment in related industries (where industries are considered related if they exceed a certain threshold). However, even in these regressions, skill-related employment tends to remain highly significant.\(^{31}\) Overall, we therefore conclude that the predictive validity of skill-relatedness is strong vis-à-vis co-location-based and input-output relatedness:

**Stylized fact 6:** Compared to input-output and co-location-based relatedness indices, skill relatedness is a relatively strong predictor of industries’ regional growth and entry rates.

---

\(^{30}\) This is not due to high correlations among inter-industry relatedness measures. Indeed, at the 2-digit level, the correlation is 0.50 between \(SR_{ij}\) and \(IO_{ij}\), 0.36 between \(SR_{ij}\) and \(CL_{ij}\), 0.36, and 0.40 between \(IO_{ij}\) and \(CL_{ij}\). At the 4-digit level, these correlations drop to 0.29, 0.31 and 0.19, respectively. However, at the industry-region level, correlations among \(\log(E_{iirt}^{SR})\), \(\log(E_{iirt}^{IO})\) and \(\log(E_{iirt}^{CL})\) run from a maximum of 0.98 at the 2-digit level to 0.94 at the 4-digit level. Indeed, at the 2-digit level, where correlations are highest and the number of observations is lowest, multicollinearity issues become unsurmountable.

\(^{31}\) Results are available upon request.
Skill relatedness and reallocation frictions

One way of thinking of the skill-relatedness patterns documented so far is that they represent constraints to labor mobility. Such constraints may hinder the efficient reallocation of workers from declining to expanding industries. This may affect a region’s resilience to adverse shocks. Indeed, just as a lack of geographical labor mobility makes it harder for economies to deal with asymmetric regional shocks, the constrained mobility across industries raises similar concerns. However, the degree to which such mobility constraints cause reallocation frictions depends on whether or not related industries experience correlated shocks. That is, only if industries that are skill related typically grow or shrink at the same time will it be hard to reallocate workers in skill-preserving ways.

To assess to what extent this is the case, we need to explore whether the differences in growth rates of two industries are associated with their skill relatedness. To do so, we first calculate skill relatedness among 3-digit industries, using only flows in the period 1999/00 to 2002/03. For the remaining years, 2004 to 2008, we calculate the absolute difference in growth rates for each pair of industries:

\[ G_{dif_{ij}} = \frac{E_{i,2008} - E_{i,2004}}{E_{j,2008} - E_{j,2004}} \]

When industries \( i \) and \( j \) grow at exactly the same rate, \( G_{dif_{ij}} \) is zero and \( G_{dif_{ij}} \) increases as their growth rates diverge. We find that at -0.17, the Spearman rank correlation between \( R_{ij} \) and \( G_{dif_{ij}} \) is negative yet small. Accordingly, skill-related industries tend to have only weakly similar growth rates, which should limit problems for the economy as a whole to reallocate labor from shrinking to growing industries. Indeed, in Appendix E, we show that if redundant workers are reallocated to the most skill-related industry that experiences labor shortages, this reallocation can be achieved in skill-preserving ways.

\[ \text{See for instance, Siebert (1997) and Bentivogli and Pagano (1999) for a discussion of limited labor mobility among European regions in the context of Europe’s Economic and Monetary Union as an optimal currency area.} \]
In spite of this finding, reallocation bottlenecks may still exist in individual regions. Therefore, we repeat these analyses, but define $G_{dif_{ij}}$ in terms of industries' growth rates within a planning region. Figure 3 shows the histogram for the Spearman rank correlations between these regional growth differentials and skill relatedness.

FIGURE 3 ABOUT HERE

At the regional level, reallocation problems are even less pronounced than for Germany as a whole. For some regions rank correlations are even positive and the lowest observed rank correlation in a region is -0.126. These findings suggest the following stylized fact:

**Stylized fact 7:** Skill-related industries typically do not exhibit highly similar growth rates.

Therefore, in spite of constraints to inter-industry labor flows, labor surpluses of shrinking industries can typically be absorbed by growing industries in a skill-preserving way.

5. Conclusions and future research

Our analyses of inter-industry labor flows in the German economy have yielded the following seven stylized facts. First, workers often switch jobs between industries that belong to different (highly aggregated) sections of the industry classification system (Stylized Fact 1). Although this fact may be taken as an indication that workers can change industries relatively freely and *a fortiori* that human capital cannot be particularly industry specific, a closer inspection shows that industry switches are far from random. On the contrary, most labor flows take place within a narrow set of industry pairs (Stylized Fact 2): on average, 50% of job switchers move to industries that represent just 3.4% of total employment in Germany. Furthermore, after controlling for the overall flow rates of industries, the underlying structure of inter-industry labor flows hardly changes over time (Stylized Fact 5) and does not just reflect industries’
co-location patterns (Stylized Fact 3). Moreover, job switchers with different wages and occupations follow remarkably similar flow-patterns (Stylized Fact 4).

These findings suggest that inter-industry labor flows reveal how industries are connected to one another in terms of their human-capital requirements. In line with recent literature, we therefore interpret labor flows between two industries that exceed a well-defined random benchmark as a sign that these industries are skill related. Following the literature on related diversification in economic geography, we test whether this labor-flow-based skill-relatedness measure predicts local industries’ entry and growth rates. These analyses show that the skill-relatedness index yields better growth predictions than inter-industry relatedness measures based on co-location patterns or input-output relations (Stylized Fact 6). However, although the presence of skill-related industries in a region predicts growth of local industries, this does not imply that skill-related industries exhibit correlated growth patterns. Consequently, the constraints skill relatedness impose on inter-industry labor flows do not translate into major obstacles to reallocating workers from shrinking to growing industries (Stylized Fact 7).

Overall, we believe that these findings show that inter-industry labor flows are at present understudied and merit further scholarly attention. For instance, skill-relatedness matrices can be further exploited in research where inter-industry linkages plays a role, ranging from labor economics (Poletaev and Robinson, 2008; Gathmann and Schönberg, 2010) and economic geography (Porter, 2003; Neffke et al., 2011; Boschma et al., 2013) to development economics (Hidalgo et al., 2007) and strategic management (e.g., Teece et al., 1994; Farjoun, 1994; Bryce and Winter, 2009; Lien and Klein, 2009). Moreover, skill-relatedness matrices may find various policy applications, from employment and retraining programs to cluster policy. Furthermore, we have shown that analyzing inter-industry flows offers new ways to study labor markets and the flexibility of a labor force. Finally, although skill-relatedness does not change much in the relatively short period we study, if technological progress affects
industries’ skill requirements we expect more drastic changes over longer time horizons. Labor-flow matrices may, therefore, offer new ways to analyze technological change. We hope that making the tools and skill-relatedness matrices presented in this paper available as an online appendix\textsuperscript{33} will facilitate progress on these and other topics.

\textsuperscript{33} A link to skill-relatedness estimates for all different classification systems in use in Germany between 1975 and 2014 can be found on the first author’s website.
References


Jahrbücher für Nationalökonomik and Statistik. 2016. Special Issue on 25 years of German reunification 236(2).


## Tables

**Table 1: Cross-industry labor flows by labor-market segment**

<table>
<thead>
<tr>
<th>LABOR-MARKET SEGMENT</th>
<th>WORKERS (THOUSANDS)</th>
<th>WAGES</th>
<th>GERMANY</th>
<th>GEOGRAPHY</th>
<th>OCCUPATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>high</td>
<td>low</td>
<td>all</td>
<td>East</td>
</tr>
<tr>
<td>employment</td>
<td>19,897.1</td>
<td>9,947.6</td>
<td>9,914.8</td>
<td>3,890.2</td>
<td>15,958.8</td>
</tr>
<tr>
<td>job switchers</td>
<td>1,206.7</td>
<td>554.3</td>
<td>652.4</td>
<td>194.2</td>
<td>658.7</td>
</tr>
<tr>
<td>no industry switch</td>
<td>321.1</td>
<td>165.2</td>
<td>155.9</td>
<td>56.2</td>
<td>246.1</td>
</tr>
<tr>
<td>industry switch</td>
<td>885.7</td>
<td>389.1</td>
<td>496.5</td>
<td>138.0</td>
<td>694.8</td>
</tr>
<tr>
<td>different section</td>
<td>519.8</td>
<td>215.9</td>
<td>303.8</td>
<td>82.1</td>
<td>408.0</td>
</tr>
<tr>
<td>same section</td>
<td>365.9</td>
<td>173.2</td>
<td>192.7</td>
<td>56.0</td>
<td>286.8</td>
</tr>
<tr>
<td>same sub-section</td>
<td>301.6</td>
<td>145.4</td>
<td>156.2</td>
<td>50.1</td>
<td>231.1</td>
</tr>
<tr>
<td>same 2-digit industry</td>
<td>225.9</td>
<td>109.8</td>
<td>116.0</td>
<td>40.4</td>
<td>170.4</td>
</tr>
<tr>
<td>same 3-digit industry</td>
<td>117.3</td>
<td>58.8</td>
<td>58.5</td>
<td>20.5</td>
<td>88.6</td>
</tr>
<tr>
<td>same 4-digit industry</td>
<td>62.3</td>
<td>31.7</td>
<td>30.6</td>
<td>10.4</td>
<td>47.6</td>
</tr>
</tbody>
</table>

**PERCENTAGES**

| no industry switch | 26.6% | 29.8% | 23.9% | 28.9% | 26.2% | 26.9% | 25.4% | 22.2% | 27.8% | 27.9% | 22.1% | 15.7% | 29.6% | 33.3% |
| industry switch | 73.4% | 70.2% | 76.1% | 71.1% | 73.8% | 73.1% | 74.6% | 77.8% | 72.2% | 72.1% | 77.9% | 84.3% | 70.4% | 66.7% |
| different section | 58.7% | 55.5% | 61.2% | 59.5% | 58.7% | 59.1% | 57.3% | 58.4% | 53.8% | 58.4% | 65.2% | 58.5% | 78.8% | 73.9% |
| same section | 41.3% | 44.5% | 38.8% | 40.5% | 41.3% | 40.9% | 42.7% | 41.6% | 46.2% | 41.6% | 34.8% | 41.5% | 21.2% | 26.1% |
| same sub-section | 34.1% | 37.4% | 31.4% | 36.3% | 33.3% | 33.2% | 37.2% | 34.7% | 43.1% | 38.0% | 29.8% | 38.8% | 20.2% | 25.7% |
| same 2-digit industry | 25.5% | 28.2% | 23.4% | 29.3% | 24.5% | 25.3% | 26.3% | 23.9% | 30.4% | 29.3% | 20.3% | 19.0% | 16.2% | 19.1% |
| same 3-digit industry | 13.2% | 15.1% | 11.8% | 14.9% | 12.7% | 13.0% | 14.0% | 11.3% | 13.5% | 21.7% | 9.3% | 8.7% | 4.4% | 5.6% |
| same 4-digit industry | 7.0% | 8.1% | 6.2% | 7.5% | 6.8% | 6.9% | 7.7% | 5.7% | 7.1% | 10.9% | 4.4% | 6.8% | 3.0% | 3.3% |

The table contains information about absolute and relative employment levels, jobs switching and cross-industry flows in each labor-market segment averaged over 1999-2007. In the section WAGES, column <all> refers to all workers, column <high> refers to workers with wages above their industry’s median, column <low> to workers below this median. Long-distance flows are labor flows between establishments that are at least 100 km apart. The occupations are as defined in Table C1.
Table 2: $FLEX(0.50)$ estimates by labor-market segment

<table>
<thead>
<tr>
<th>Segment</th>
<th>$FLEX(50)$ (inc. intra-ind flows)</th>
<th>$FLEX(50)$ (exc. intra-ind flows)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>3.4% (33.2%)</td>
<td>5.9% (33.0%)</td>
</tr>
<tr>
<td>high wage</td>
<td>2.6% (22.8%)</td>
<td>4.2% (22.6%)</td>
</tr>
<tr>
<td>low wage</td>
<td>3.4% (31.6%)</td>
<td>5.9% (31.4%)</td>
</tr>
<tr>
<td>East / West Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>4.0% (29.1%)</td>
<td>6.4% (28.9%)</td>
</tr>
<tr>
<td>West</td>
<td>2.2% (28.0%)</td>
<td>3.9% (27.7%)</td>
</tr>
<tr>
<td>Geography of flows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>local</td>
<td>3.4% (31.8%)</td>
<td>6.0% (31.6%)</td>
</tr>
<tr>
<td>long-distance</td>
<td>2.3% (22.3%)</td>
<td>3.4% (22.1%)</td>
</tr>
<tr>
<td>Occupations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>managers</td>
<td>2.3% (8.5%)</td>
<td>2.9% (8.4%)</td>
</tr>
<tr>
<td>sales</td>
<td>3.2% (25.2%)</td>
<td>4.6% (24.5%)</td>
</tr>
<tr>
<td>accountants</td>
<td>3.2% (25.2%)</td>
<td>4.6% (24.5%)</td>
</tr>
<tr>
<td>office clerks</td>
<td>3.8% (18.8%)</td>
<td>15.0% (22.8%)</td>
</tr>
<tr>
<td>IT</td>
<td>16.8% (26.0%)</td>
<td>5.6% (18.4%)</td>
</tr>
<tr>
<td>cleaners</td>
<td>16.5% (30.3%)</td>
<td>6.2% (18.4%)</td>
</tr>
<tr>
<td>security</td>
<td>11.9% (22.0%)</td>
<td>8.9% (16.7%)</td>
</tr>
</tbody>
</table>

$FLEX(50)$ equals the employment weighted average employment share (across industries of origin) of the 5-digit destination industries with the highest labor-flow-to-employment ratios that taken together absorb 50% of all labor flows originating from a given 5-digit industry. All values represent averages across the period 1999-2007. The number provided in parenthesis represents the $FLEX(50)$ that was simulated under the null-model.
Table 3: Correlations of skill-relatedness estimates of local versus long-distance flows

<table>
<thead>
<tr>
<th></th>
<th>local</th>
<th>long-distance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>own correlation</strong></td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>local</strong></td>
<td>n.a.</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>long-distance</strong></td>
<td>0.87</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The table presents correlations between the skill-relatedness \( \bar{R}_{ij} \) matrices for the combination of labor-market segments given in the rows and columns. The first row of the table provides the average correlation between skill-relatedness matrices for two consecutive years of the labor-market segment in the columns. In the remaining cells, the top row represents the bias-corrected correlation using equation (3) and the bottom row the correlation between relatedness matrices averaged over all available years. Industries have been aggregated to the 3-digit level. Local flows are flows taking place over a road-distance of below 100 km, flows beyond this distance are labelled long-distance.
<table>
<thead>
<tr>
<th></th>
<th>high East</th>
<th>low East</th>
<th>high West</th>
<th>low West</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>own correlation</strong></td>
<td>0.45</td>
<td>0.43</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>high East</strong></td>
<td>n.a.</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>low East</strong></td>
<td>0.94</td>
<td>n.a.</td>
<td>0.74</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>high West</strong></td>
<td>0.79</td>
<td>0.74</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td><strong>low West</strong></td>
<td>0.75</td>
<td>0.79</td>
<td>0.93</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Idem Table 3. High (low) East (West) represents high-wage (low-wage) workers in eastern (western) Germany.
Table 5: Correlations of skill-relatedness estimates for different occupations

<table>
<thead>
<tr>
<th></th>
<th>managers</th>
<th>sales</th>
<th>accountants</th>
<th>office clerks</th>
<th>IT</th>
<th>security</th>
<th>cleaners</th>
</tr>
</thead>
<tbody>
<tr>
<td>own correlation</td>
<td>0.42</td>
<td>0.43</td>
<td>0.29</td>
<td>0.44</td>
<td>0.33</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>managers</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>sales</td>
<td>0.92</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>accountants</td>
<td>0.86</td>
<td>0.80</td>
<td>0.80</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>office clerks</td>
<td>0.91</td>
<td>0.89</td>
<td>0.78</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>IT</td>
<td>0.83</td>
<td>0.78</td>
<td>0.89</td>
<td>0.76</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>security</td>
<td>0.44</td>
<td>0.38</td>
<td>0.53</td>
<td>0.39</td>
<td>0.44</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>cleaners</td>
<td>0.50</td>
<td>0.41</td>
<td>0.60</td>
<td>0.47</td>
<td>0.57</td>
<td>0.79</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Idem Table 3. Row and column labels refer to broad occupational groupings.
Table 6: Changes in skill relatedness between 1999 and 2008

<table>
<thead>
<tr>
<th>segment</th>
<th>stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td></td>
</tr>
<tr>
<td>all wages</td>
<td>0.93</td>
</tr>
<tr>
<td>high wages</td>
<td>0.93</td>
</tr>
<tr>
<td>low wages</td>
<td>0.93</td>
</tr>
<tr>
<td>East / West Germany</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>0.90</td>
</tr>
<tr>
<td>West</td>
<td>0.93</td>
</tr>
<tr>
<td>Geography of flows</td>
<td></td>
</tr>
<tr>
<td>local</td>
<td>0.93</td>
</tr>
<tr>
<td>long-distance</td>
<td>0.95</td>
</tr>
<tr>
<td>Occupations</td>
<td></td>
</tr>
<tr>
<td>managers</td>
<td>0.94</td>
</tr>
<tr>
<td>sales</td>
<td>0.97</td>
</tr>
<tr>
<td>accountants</td>
<td>0.92</td>
</tr>
<tr>
<td>office clerks</td>
<td>0.95</td>
</tr>
<tr>
<td>IT</td>
<td>0.93</td>
</tr>
<tr>
<td>cleaners</td>
<td>0.91</td>
</tr>
<tr>
<td>security</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Stability is the attenuation-bias-corrected correlation between skill relatedness in year 1999 and 2007 of a given labor-market segment. Measurement error variances are assessed by the correlations between skill-relatedness estimates for 1999/00 and 2000/01 and for 2006/07 and 2007/08 of the corresponding labor-market segment.
Table 7a: Local industry growth and entry regressions (3-digit industries)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Eirt)</td>
<td>-0.035*** (0.0036)</td>
<td>-0.030*** (0.0015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Eirt_SR)</td>
<td>0.171*** (0.0416)</td>
<td>0.158*** (0.0169)</td>
<td>0.074* (0.0399)</td>
<td>-0.061 (0.0705)</td>
</tr>
<tr>
<td>log(Eirt_CL)</td>
<td>0.126 (0.1911)</td>
<td>0.082 (0.0854)</td>
<td>-0.152 (0.5325)</td>
<td>1.448* (0.8588)</td>
</tr>
<tr>
<td>log(Eirt_IO)</td>
<td>0.004 (0.0234)</td>
<td>-0.016 (0.0100)</td>
<td>-0.003 (0.0264)</td>
<td>-0.077 (0.0538)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.054</td>
<td>0.175</td>
<td>0.226</td>
<td>0.253</td>
</tr>
<tr>
<td>N</td>
<td>12,408</td>
<td>12,168</td>
<td>2,046</td>
<td>2,046</td>
</tr>
</tbody>
</table>

***: p<0.01, **: p<0.05, *: p<0.10, robust standard errors in parentheses. All models include region and industry fixed effects. Regional units are defined as Germany’s planning regions and industries are aggregated at the 3-digit level, excluding industries in agriculture, fishing and mining (NACE 01-14), non-traded industries (NACE 40-59) and public sector industries (NACE 75-99). Columns (1) and (2) report OLS regressions of the logarithm of annualized growth rates over a one- and over a five-year period. Columns (3) and (4) report the outcomes of OLS regressions with as a dependent variable a dummy variable that evaluates whether or not a new local industry enters a region within a one- or a five-year period.

Table 7b: Local industry growth and entry regressions (4-digit industries)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Eirt)</td>
<td>-0.043*** (0.0028)</td>
<td>-0.033*** (0.0011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Eirt_SR)</td>
<td>0.200*** (0.0333)</td>
<td>0.146*** (0.0120)</td>
<td>0.051*** (0.0140)</td>
<td>0.104*** (0.0242)</td>
</tr>
<tr>
<td>log(Eirt_CL)</td>
<td>-0.041 (0.2279)</td>
<td>0.035 (0.0826)</td>
<td>0.067 (0.2409)</td>
<td>-0.195 (0.3628)</td>
</tr>
<tr>
<td>log(Eirt_IO)</td>
<td>0.042** (0.0192)</td>
<td>0.005 (0.0077)</td>
<td>0.015 (0.0127)</td>
<td>-0.019 (0.0191)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.082</td>
<td>0.201</td>
<td>0.153</td>
<td>0.207</td>
</tr>
<tr>
<td>N</td>
<td>21,855</td>
<td>21,050</td>
<td>9,309</td>
<td>9,309</td>
</tr>
</tbody>
</table>

Idem Table 7a, but using industries aggregated at the 4-digit level.
Figures

Figure 1: Flexibility of the German labor force

The vertical axis shows the percentage of employment represented by industries that together absorb the percentage of industry switchers depicted on the horizontal axis. Values include within-industry flows, are averaged across the years 1999-2007 and are calculated for the German labor market as a whole. The dotted line provides a random benchmark. The thin grey line represents the 45° as a point of reference.
Figure 2a: Clustering of inter-industry linkages in Germany (1999-2008)

The shading in the figure represents the average symmetrized skill-relatedness matrix for Germany at the 5-digit level using all yearly flows between 1999 and 2008. Rows and columns are sorted according to an average-linkages hierarchical clustering algorithm.
Figure 2b: Network depiction of inter-industry linkages in Germany (1999-2008)

The network depicts the strongest 651 links among 3-digit industries in the symmetrized average skill-relatedness matrix for Germany as a whole, using all yearly flows between 1999 and 2008. The layout is based on the organic layout procedure in the Cytoscape software, manually adjusted to increase the clarity of the graph by minimizing edge crossings. Labels are omitted for small industries. Color codes represent subsections in the NACE classification.
Figure 3: Regional labor reallocation: correlation($R_{ij}$,$G_{dif}$) by region

The graph shows the histogram of region-specific rank correlations between skill relatedness and the absolute difference in local growth rates within an industry pair.
Appendix A: Classification systems

The German industry classification system has seen three major overhauls since 1975. From 1975 to 1998 the Statistical Classification of Economic Activities in Germany 1973 (WZ 1973) was used. In 1999 the WZ 1993 was implemented and used until 2003, followed by WZ 2003 (used from 2003-2008) and the WZ 2008 in 2008. The WZ 1993, WZ 2003 and WZ 2008 were harmonized at the 4-digit level with the European NACE (Nomenclature générale des Activités économiques dans les Communautés Européennes) 1.0, NACE Revision 1.1 and NACE Revision 2.0 classifications. The classification systems in the period of our study, WZ 1993 and WZ 2003, consist of six levels. Industries are first grouped into sets coded by two capital letters, the first of which denotes the section to which the industry belongs and the second divides these sections into sub-sections. The remaining four levels are indicated by the first 2, 3, 4 and 5 digits of a numeric code.

With the introduction of the WZ 2003 a minor reclassification took place. Therefore we harmonize the WZ 1993 and WZ 2003, taking the WZ 2003 as a starting point. For plants that exist before 2003, we apply the WZ 2003 industry codes to earlier years, provided that the establishment’s WZ 1993 code remains the same. Next, we construct a concordance between WZ 1993 and WZ 2003 to fill in the gaps that occur when plants exit or change their WZ 1993 codes before 2003. This concordance is based on information from the year 2003 in which establishments’ industries are recorded in both classification systems. For most 5-digit WZ 1993 industry-codes, we find all but unique translations to the WZ 2003 system. However, in some cases single WZ 1993 industries were split or multiple WZ 1993 were merged, we merge 60 out of 1043 WZ 2003 classes into 29 combined industries. The exact procedures are coded in the STATA do-files accompanying this paper.
Appendix B: Identifying spurious labor flows

Hethey and Schmieder (2010) argue that the reliability of the establishment identifiers in the HES data can be improved by analyzing labor flows among establishments. For instance, if all workers in a new establishment come from the same other establishment, this establishment is likely to be a spinoff or an existing establishment with a changed code. To separate real labor flows from the spurious flows that arise from recoding establishments, we analyze inter-establishment labor flows. Whether a flow should be considered spurious depends on the size of the establishments of the flow’s origin and destination.

For establishments with fewer than five employees, a flow is considered spurious if all employees either came from or went to the same establishment (the exception are establishments with only one employee). For plants with five or more employees, flows are considered spurious if they represent at least 30% of the employment in the establishment where the flow originates or ends. Furthermore, flows of 100 employees or more are considered spurious regardless of establishment sizes. We illustrate these choices in Figures B1 and B2, which depict the yearly average number of establishment-to-establishment transitions that are excluded at different thresholds.

Figure B1 shows how many individuals in establishments with at least 5 employees are dropped. A threshold of 30% (the lower bound used by Hethey and Schmieder, 2010) drops about 27.5% of the about 1.8 million switchers. This percentage does not change much until the cutoff exceeds 60%. Figure B2 plots the number of individuals dropped for smaller establishments. Because small establishments represent only a modest number of job switchers, not many flows are lost regardless of the chosen threshold. The imposition of a 100% threshold in this paper implies that about 32,200 establishment switchers are disregarded per year. The final restriction we impose is that inter-establishment flows should never consists of blocks larger than 100 individuals. This condition reduces the sample by another 11,100 individuals per year.
Figure B1: Suspect flows (large establishments)

Number of job switchers that are excluded in establishments with 5 or more employees at different thresholds.

Figure B2: Suspect flows (small establishments)

Bars represent the number of switchers that are excluded at the threshold depicted above each bar.
# Appendix C: Occupational groupings

Table C: Number of employees by selected occupations and occupational grouping

<table>
<thead>
<tr>
<th>Occupations</th>
<th>average # emp. (1999-2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Management Occupations</strong></td>
<td></td>
</tr>
<tr>
<td>751: Entrepreneurs, managing directors, divisional managers</td>
<td>341,450</td>
</tr>
<tr>
<td>762: Executive administrative professionals</td>
<td>100,476</td>
</tr>
<tr>
<td><strong>Sales Occupations</strong></td>
<td><strong>1,448,132</strong></td>
</tr>
<tr>
<td>681: Wholesale and retail trade buyers, buyers</td>
<td>367,342</td>
</tr>
<tr>
<td>682: Salespersons</td>
<td>772,100</td>
</tr>
<tr>
<td>687: Commercial agents, travelers</td>
<td>175,604</td>
</tr>
<tr>
<td>701: Forwarding business dealers</td>
<td>73,839</td>
</tr>
<tr>
<td>703: Publicity occupations</td>
<td>59,247</td>
</tr>
<tr>
<td><strong>Accountants</strong></td>
<td><strong>301,287</strong></td>
</tr>
<tr>
<td>753: Chartered accountants, tax advisers</td>
<td>114,940</td>
</tr>
<tr>
<td>771: Cost accountants, valuers</td>
<td>37,693</td>
</tr>
<tr>
<td>772: Accountants</td>
<td>148,654</td>
</tr>
<tr>
<td><strong>Office clerks</strong></td>
<td><strong>2,653,181</strong></td>
</tr>
<tr>
<td>781: Office specialists</td>
<td>2,653,181</td>
</tr>
<tr>
<td><strong>IT specialists</strong></td>
<td><strong>399,136</strong></td>
</tr>
<tr>
<td>774: Data processing specialists</td>
<td>399,136</td>
</tr>
<tr>
<td><strong>Cleaners</strong></td>
<td><strong>118,921</strong></td>
</tr>
<tr>
<td>793: Doormen, caretakers</td>
<td>145,238</td>
</tr>
<tr>
<td>933: Household cleaners</td>
<td>193,222</td>
</tr>
<tr>
<td>934: Glass, buildings cleaners</td>
<td>38,508</td>
</tr>
<tr>
<td><strong>Security personnel</strong></td>
<td><strong>231,730</strong></td>
</tr>
<tr>
<td>791: Factory guards, detectives</td>
<td>17,977</td>
</tr>
<tr>
<td>792: Watchmen, custodians</td>
<td>100,944</td>
</tr>
</tbody>
</table>

Occupations are classified in 3-digit codes by the German Classifications of Occupations 1973.
Appendix D: Derivation attenuation bias correction

In this appendix, we derive the bias-correction equation of equation (3). First, \( r_{t,s} \) as the stacked the \( R \)-matrix of a labor-market segment \( s \) in year \( t \) into a vector, omitting diagonal elements:

\[
r_{t,s} = \begin{pmatrix} 
R_{1,2}^{t,s} & \vdots \\
\vdots & \ddots & \vdots \\
R_{1,n}^{t,s} & \vdots & R_{n,1}^{t,s} \\
R_{n,(n-1)}^{t,s} & \end{pmatrix}
\]

In this vector, \( R_{i,j}^{t,s} \) represents the skill relatedness between industries \( i \) and \( j \) in year \( t \) and segment \( s \). As described in equation (2), relatedness estimates can be thought of as consisting of two components: a structural, yet unobserved component and an orthogonal random component.

\[
\hat{r}_{st} = r_s + \varepsilon_{st}
\]

where \( \hat{r}_{st} \) represents the observed vector of relatedness estimates, \( r_s \) the true relatedness, and \( \varepsilon_{st} \) a measurement error. Note that \( r_s \) does not depend on \( t \). That is, we assume that the true underlying relatedness is unchanging.

For notational clarity, we denote the estimated skill-relatedness vector in segment \( s \) in year \( t \), \( \hat{r}_{st} \), and its true vector, \( r_s \), by \( \hat{x}_t \) and \( x \). Similarly, \( \hat{y}_t \) and \( y_t \) correspond to the estimated and real skill-relatedness vectors in segment \( s' \). The correlation between the skill-relatedness estimates of the two segments can now be written as:

\[
\text{Corr}[\hat{x}_t, \hat{y}_t] = \frac{\text{Cov}[x + \varepsilon_x, y + \varepsilon_y]}{\sqrt{\text{Var}[x + \varepsilon_x] \text{Var}[y + \varepsilon_y]}}
\] (D1)

Or,

\[
\text{Corr}[\hat{x}_t, \hat{y}_t] = \frac{\text{Cov}[x, y] + \text{Cov}[x, \varepsilon_x] + \text{Cov}[\varepsilon_x, y] + \text{Cov}[\varepsilon_x, \varepsilon_y]}{(\text{Var}[x] + 2\text{Cov}[x, \varepsilon_x] + \text{Var}[\varepsilon_x])(\text{Var}[y] + 2\text{Cov}[y, \varepsilon_y] + \text{Var}[\varepsilon_y])}
\]
Let us now assume that the measurement errors are uncorrelated with the true skill-relatedness values:

Assumption 1a: $\text{Corr}(x, \varepsilon_{xt}) = \text{Corr}(y, \varepsilon_{yt}) = 0$

Assumption 1b: $\text{Corr}(x, \varepsilon_{yt}) = \text{Corr}(y, \varepsilon_{xt}) = 0$

Because skills are latent constructs, Assumption 1a can be interpreted as a definition: whatever it is that we will refer to as skill relatedness, its estimated value can be decomposed into an invariant, structural term and into an error term. Let us further assume that the error terms for both relatedness types are uncorrelated as well:

Assumption 2: $\text{Corr}(\varepsilon_{xt}, \varepsilon_{yt}) = 0$

Using the assumptions, we can rewrite (D1) as:

$$\text{Corr}[\hat{x}_t, \hat{y}_t] = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]+\text{Var}[\varepsilon_{xt}])(\text{Var}[y]+\text{Var}[\varepsilon_{yt}]})}$$

or, multiplying by $\sqrt{\text{Var}[x]/\text{Var}[y]}$, $\sqrt{\text{Var}[x]/\text{Var}[y]}$:

$$\text{Corr}[\hat{x}_t, \hat{y}_t] = \frac{\text{Cov}[x,y]}{\text{Var}[x] \sqrt{\text{Var}[y]}}$$

$$\text{Cov}[x,y] \sqrt{\text{Var}[x] + \text{Var}[\varepsilon_{xt}]} \sqrt{\text{Var}[y] + \text{Var}[\varepsilon_{yt}]}$$

which is equation (9) in the main text. Rearranging terms and using the fact that $\rho_{xy} = \text{Corr}(x,y) = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x] \text{Var}[y]}}$ yields the following expression for the real correlation across segments:

$$\rho_{xy} = \rho_{\varepsilon_{xt} \varepsilon_{yt}} \sqrt{\frac{\text{Var}[x] + \text{Var}[\varepsilon_{xt}]}{\text{Var}[x]}} \sqrt{\frac{\text{Var}[y] + \text{Var}[\varepsilon_{yt}]}{\text{Var}[y]}} = \rho_{\varepsilon_{xt} \varepsilon_{yt}} \sqrt{1 + \frac{\text{Var}[\varepsilon_{xt}]}{\text{Var}[x]} \sqrt{1 + \frac{\text{Var}[\varepsilon_{yt}]}{\text{Var}[y]}}}$$ (D2)

(D2) shows that the downward bias in the measured correlation depends on the relative sizes of the error variances compared to the variance of the relatedness types. Therefore, we need an estimate of the relative size of the measurement errors. To arrive at such an estimate, we will assume that the error terms are uncorrelated over time.
Assumption 3a: $\forall t, t': t \neq t' \rightarrow Corr(\varepsilon_{xt}, \varepsilon_{xt'}) = 0$

Assumption 3b: $\forall t, t': t \neq t' \rightarrow Corr(\varepsilon_{yt}, \varepsilon_{yt'}) = 0$

The correlation between two measurements of the same skill-relatedness type can be written as:

$$Corr[\hat{x}_t, \hat{x}_{tt'}] = \frac{\text{cov}[x+\varepsilon_{xt} \text{, } x+\varepsilon_{xt'}]}{\sqrt{\text{var}[x+\varepsilon_{xt}]\text{var}[x+\varepsilon_{xt'}]}}$$

$$Corr[\hat{x}_t, \hat{x}_{tt'}] = \frac{\text{cov}[x,x]+\text{cov}[x,\varepsilon_{xt}]+\text{cov}[\varepsilon_{xt},x]+\text{cov}[\varepsilon_{xt},\varepsilon_{xt'}]}{\sqrt{\text{var}[x]+2\text{cov}[x,\varepsilon_{xt}]+\text{var}[\varepsilon_{xt}]}(\text{var}[x]+2\text{cov}[x,\varepsilon_{xt}]+\text{var}[\varepsilon_{xt}])}$$

Using Assumptions 3a and 3b, we arrive at:

$$Corr[\hat{x}_t, \hat{x}_{tt'}] = \frac{\text{var}[x]}{\sqrt{\text{var}[x]+\text{var}[\varepsilon_{xt}]})^2}$$

We will also assume that the measurement error has about the same variance in different years:

Assumption 4a: $\forall t, t': \text{var}(\varepsilon_{xt}) = \text{var}(\varepsilon_{xt'}) = \text{var}(\varepsilon_{x})$

Assumption 4b: $\forall t, t': \text{var}(\varepsilon_{yt}) = \text{var}(\varepsilon_{yt'}) = \text{var}(\varepsilon_{y})$

$$Corr[\hat{x}_t, \hat{x}_{tt'}] = \frac{\text{var}[x]}{\text{var}[x]+\text{var}[\varepsilon_{xt}]} \tag{D3}$$

Similar derivations for $y$ yield:

$$Corr[\hat{y}_t, \hat{y}_{tt'}] = \frac{\text{var}[y]}{\text{var}[y]+\text{var}[\varepsilon_{yt}]} \tag{D4}$$

Denoting correlation by $\rho$, we arrive at (3) by substituting (D3) and (D4) into (D2):

$$\rho_{xy} = \rho_{\hat{x}_t\hat{y}_t} \frac{1}{\sqrt{\rho_{\hat{x}_t\hat{x}_{tt'}}}} \frac{1}{\sqrt{\rho_{\hat{y}_t\hat{y}_{tt'}}}} = \frac{\rho_{\hat{x}_t\hat{y}_t}}{\sqrt{\rho_{\hat{x}_t\hat{x}_{tt'}}\rho_{\hat{y}_t\hat{y}_{tt'}}}}$$

The impact of measurement error can also be reduced by first averaging relatedness estimates over several years and then calculating correlations among these averaged values. Because we only have a finite number of years of observations, some attenuation bias will remain. This method, therefore,
provides a lower bound for the true correlations. A third approach combines both corrections. It first averages out part of the measurement error and then corrects for any remaining bias.

Let \( \bar{x} = \frac{1}{T} \sum_t \hat{x}_t \) and \( \bar{y} = \frac{1}{T} \sum_t \hat{y}_t \) be the average across \( T \) years of observations of the skill-relatedness types \( x \) and \( y \). The correlation between these two averages can be written as:

\[
\text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[\frac{1}{T} \sum (x + \varepsilon_{xt}), \frac{1}{T} \sum (y + \varepsilon_{yt})]}{\sqrt{\text{Var}[\frac{1}{T} \sum x + \varepsilon_{xt}] \text{Var}[\frac{1}{T} \sum y + \varepsilon_{yt}]}}, \tag{D5}
\]

Assuming constant variances for the measurement errors (Assumptions 4a and 4b), the numerator of (D5) can be written as:

\[
\text{Cov}[\frac{1}{T} \sum (x + \varepsilon_{xt}), \frac{1}{T} \sum (y + \varepsilon_{yt})] = T^2 \text{Cov}[x, y] + T \sum \text{Cov}[x, \varepsilon_{yt}] + T \sum \text{Cov}[\varepsilon_{xt}, y] + \sum \sum \text{Cov}[\varepsilon_{xt}, \varepsilon_{yt}], \tag{D6}
\]

By assumptions (1a), (1b), (2), (3a), (3b), (4a) and (4b), this simplifies to:

\[
\text{Cov}[\sum (x + \varepsilon_{xt}), \sum (y + \varepsilon_{yt})] = T^2 \text{Cov}[x, y] \tag{D6}
\]

Expanding the first term of the denominator of (D5), we derive the following:

\[
\text{Var}[\sum (x + \varepsilon_{xt})] = T^2 \text{Var}[x] + T \text{Var}[\varepsilon_{xt}] + \sum_t \sum \text{Var}[\varepsilon_{xt}, \varepsilon_{xt}] + \sum \sum \text{Cov}[x, \varepsilon_{xt}], \tag{D7}
\]

which simplifies under the abovementioned assumption to:

\[
\text{Var}[\sum (x + \varepsilon_{xt})] = T^2 \text{Var}[x] + T \text{Var}[\varepsilon_{xt}] \tag{D7}
\]

Due to similar considerations for \( y \), the second term in the denominator of (D5) is:

\[
\text{Var}[\sum (x + \varepsilon_{yt})] = T^2 \text{Var}[y] + T \text{Var}[\varepsilon_{yt}] \tag{D8}
\]

Substituting (D6), (D7) and (D8) into (D5)(D4), we get:
\[ \text{Corr}[\bar{x}, \bar{y}] = \frac{T^2 \text{Cov}[x,y]}{\sqrt{T^2 \text{Var}[x] + TVar[\epsilon_x]} \sqrt{T^2 \text{Var}[y] + TVar[\epsilon_y]}} = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x] + \frac{1}{T} \text{Var}[\epsilon_x]} \sqrt{\text{Var}[y] + \frac{1}{T} \text{Var}[\epsilon_y]}} \]  

(D9)

Rearranging (D4) yields the following expression for \( \text{Var}[\epsilon_x] \):

\[ \text{Var}[\epsilon_x] = \frac{\text{Var}[x]}{\text{Corr}[\bar{x}, \bar{x}_t]} - \text{Var}[x] = \frac{1 - \text{Corr}[\bar{x}_t, \bar{x}_t]}{\text{Corr}[\bar{x}_t, \bar{x}_t]} \text{Var}[x] \]  

(D10)

Substituting (D10) and its counterpart for \( y \) into (D9) yields:

\[ \text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x] + \frac{1}{T} \text{Corr}[\bar{x}_t, \bar{x}_t] \text{Var}[x]} \sqrt{\text{Var}[y] + \frac{1}{T} \text{Corr}[\bar{y}_t, \bar{y}_t] \text{Var}[y]}} \]

Now we can rearrange the terms to arrive at:

\[ \text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x] \text{Var}[y]} \sqrt{1 + \frac{1}{T} \text{Corr}[\bar{x}_t, \bar{x}_t] \text{Var}[x]} \sqrt{1 + \frac{1}{T} \text{Corr}[\bar{y}_t, \bar{y}_t] \text{Var}[y]}} \]

which can be rewritten as:

\[ \rho_{xy} = \rho_{xy} \sqrt{1 + \frac{1}{T} \frac{1 - \rho_{xy} \rho_{x_t, x'_t}}{\rho_{x_t, x'_t}}} \sqrt{1 + \frac{1}{T} \frac{1 - \rho_{y_t y'_t}}{\rho_{y_t y'_t}}} \]  

(D11)

Equation (D11), which was first proposed by Spearman (1910), and equation (3) in the main text will give the same results (in expected terms) as long as none of the assumptions are violated. This provides a convenient check for our empirical results. To test this, Tables D1 to D3 replicate Tables 4 to 6. Instead of reporting the correlation between time-averaged skill-relatedness matrices in the bottom rows of the cells, we now adjust these correlations using equation (D11). The top row in each cell is still based on the bias-correction of equation (3). In almost all cases, the bias-corrections in (D11) and (3) give almost identical results, adding credence to the bias-corrected estimates we report in the main text.
Table D1: Comparison bias-correction methods: local versus long-distance flows

<table>
<thead>
<tr>
<th></th>
<th>local</th>
<th>long-distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>own correlation</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td>local</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>long-distance</td>
<td>0.87</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>0.88</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Idem Table 4. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).

Table D2: Comparison bias-correction methods: Wages

<table>
<thead>
<tr>
<th></th>
<th>high East</th>
<th>low East</th>
<th>high West</th>
<th>low West</th>
</tr>
</thead>
<tbody>
<tr>
<td>own correlation</td>
<td>0.45</td>
<td>0.43</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td>high East</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low East</td>
<td>0.94</td>
<td>n.a.</td>
<td>0.79</td>
<td>n.a.</td>
</tr>
<tr>
<td>high West</td>
<td>0.79</td>
<td>0.74</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>low West</td>
<td>0.75</td>
<td>0.79</td>
<td>0.93</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Idem Table 5. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).

Table D3: Comparison bias-correction methods: Occupations

<table>
<thead>
<tr>
<th></th>
<th>managers</th>
<th>sales</th>
<th>accountants</th>
<th>office clerks</th>
<th>IT</th>
<th>security</th>
<th>cleaners</th>
</tr>
</thead>
<tbody>
<tr>
<td>own correlation</td>
<td>0.42</td>
<td>0.43</td>
<td>0.29</td>
<td>0.44</td>
<td>0.33</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>managers</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>sales</td>
<td>0.92</td>
<td>n.a.</td>
<td>0.89</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>accountants</td>
<td>0.86</td>
<td>0.80</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>office clerks</td>
<td>0.91</td>
<td>0.89</td>
<td>0.78</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>IT</td>
<td>0.83</td>
<td>0.78</td>
<td>0.89</td>
<td>0.76</td>
<td>0.74</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>security</td>
<td>0.44</td>
<td>0.38</td>
<td>0.53</td>
<td>0.39</td>
<td>0.44</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>cleaners</td>
<td>0.50</td>
<td>0.41</td>
<td>0.60</td>
<td>0.47</td>
<td>0.57</td>
<td>0.79</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Idem Table 6. Top rows in cells use bias-correction of equation (3), bottom rows of equation (D11).
Appendix E: Reallocation frictions

In this appendix we explore to what extent workers can be reallocated from shrinking to growing industries in skill-preserving ways. We define the reallocation potential from a shrinking industry $i$ to a growing industry $j$ as the minimum of the number of workers made redundant in the shrinking industry and the employment expansion in the growing industry. In particular, if $D_i$ is the (potentially negative) net labor demand that industry $i$ accumulates between 2004 and 2008, i.e., its employment growth between 2004 and 2008, the reallocation potential from industry $i$ to industry $j$ is:

$$POT_{ij} = \min(-D_i, D_j) \text{ if } D_i < 0 \text{ and } D_j > 0$$

$$= 0 \text{ otherwise}$$

The reallocation problem the economy needs to resolve is to create net labor flows such that the labor demand of growing industries equals the labor surplus of shrinking industries. Ideally, this reallocation would shift workers between highly skill-related industries. We now explore at which degree of skill relatedness workers could have been reallocated, had labor shortages been resolved “optimally”, that is, by reallocating labor in decreasing order of skill relatedness. To disregard overall growth, we first scale shortages by a common factor such that overall shortages and surpluses sum to zero. Next, we move down the list of industry pairs in descending order of skill relatedness, shifting workers from shrinking to growing industries until all labor surpluses are exhausted. The average $\bar{R}_{ij}$ of these

---

34 In this context, the correlation between reallocation potential and skill relatedness is more important than the correlation between growth rates reported in the main text. The reason is that reallocation potentials reflect the size of the labor surplus that needs to be reallocated, whereas the difference in growth rates does not take industries’ absolute sizes into account. Positive correlations imply that reallocation potentials are larger for more skill-related industries. Indeed, none of the German spatial planning regions displays significantly negative correlations between industries’ skill relatedness and their local reallocation potential.

35 In principle, trading off lower skill relatedness in some industry pairs against higher skill relatedness elsewhere might lead to even higher average levels of skill relatedness. However, to make such trade-offs, skill relatedness needs to have a clear ratio-scale interpretation, which is not guaranteed.

36 Essentially, this assumes that all growing industries meet their increased labor demand to the same extent with new workers and that the downsizing in all shrinking industries relies to the same extent on workers exiting the labor market.
artificial reallocation flows is 0.415, substantially higher than the average $\bar{R}_{ij}$ for actual job switches (0.081) in the same period, suggesting that industries’ changing labor demands could have been easily met in skill-preserving ways.

To identify potential frictions at the level of local labor markets, we repeat the reallocation thought-experiment for Germany’s 96 spatial planning regions (i.e., Raumordnungsregionen). Figure E1 shows that, in all regions, the average $\bar{R}_{ij}$ of artificial reallocation flows exceeds the average $\bar{R}_{ij}$ for actual job switches by far, reiterating the conclusion in Stylized Fact 7 that, also at the regional level, the constraints to inter-industry mobility do not need to negatively impact the efficient reallocation of labor resources.

Figure E1: Regional labor reallocation: average skill relatedness by region

The graph shows the histogram of the average skill relatedness at which workers can be reallocated from shrinking to growing industries in regions.