Abstract
Neffke and Henning propose the concept of skill-relatedness to measure the degree to which sets of industries have similar human capital requirements. The skill-relatedness among industries is inferred from cross-industry labor flows, relying on the fact that individuals are likely to switch to jobs in industries that value their previously acquired work-related skills to avoid rendering large parts of their human capital redundant. We set out to use this methodology to construct a range of skill-relatedness types to shed light on the different ways in which industries are connected to one another. For instance, we expected that industries that are similar in the skills required for their management tasks might be dissimilar in terms of the skills their sales channels required. This would allow us to distinguish between sales and management skill-relatedness. We estimate a range of skill-relatedness types by splitting up all labor flows in the
German labor market, using four different segmentations: a segmentation by wage levels, by occupational groupings, by the geographic distance involved in the flows, and by the distinction between East and West German labor markets. Surprisingly however, we find that the same general skill-relatedness guides labor flows in all different labor market segments. What is more, this relatedness also predicts the direction of diversification in a sample of Swedish firms surprisingly well. These findings suggest that the skill-relatedness structure has universal characteristics, affecting a wide range of dissimilar economic actors.
Spot the differences!

The invariable nature of skill-relatedness in Germany
Abstract

Neffke and Henning propose the concept of skill-relatedness to measure the degree to which sets of industries have similar human capital requirements. The skill-relatedness among industries is inferred from cross-industry labor flows, relying on the fact that individuals are likely to switch to jobs in industries that value their previously acquired work-related skills to avoid rendering large parts of their human capital redundant. We set out to use this methodology to construct a range of skill-relatedness types to shed light on the different ways in which industries are connected to one another. For instance, we expected that industries that are similar in the skills required for their management tasks might be dissimilar in terms of the skills their sales channels required. This would allow us to distinguish between sales and management skill-relatedness. We estimate a range of skill-relatedness types by splitting up all labor flows in the German labor market, using four different segmentations: a segmentation by wage levels, by occupational groupings, by the geographic distance involved in the flows, and by the distinction between East and West German labor markets. Surprisingly however, we find that the same general skill-relatedness guides labor flows in all different labor market segments. What is more, this relatedness also predicts the direction of diversification in a sample of Swedish firms surprisingly well. These findings suggest that the skill-relatedness structure has universal characteristics, affecting a wide range of dissimilar economic actors.
1 Introduction

Industries are often depicted as a self-contained groups of firms that compete among each other for a greater share of the same market. This perspective is useful if the goal is to understand firms’ strategic behavior in competitive markets. However, even a casual observer will note that industries are by no means self-contained. Large firms seldom operate only on a single market, innovations developed in one industry often find their ways into the products and processes of other industries, interconnected industries form geographically collocated clusters, and many individuals switch industries at some point in their careers. This suggests that the boundaries of an industry are permeable.

As a consequence of the porous nature of inter-industry boundaries, diversification is an important phenomenon in a range of economic entities. Firms diversify into new industries to expand their product ranges or to engage in higher margin activities (Teece 1982, Hoskisson and Hitt 1990, Montgomery 1994, Teece et al. 1994, Palich et al. 2000). Regions and countries climb the development ladder by setting up new export industries (Hidalgo et al. 2007, Frenken and Boschma, 2011, Neffke et al. 2011). Individuals switch to jobs that are a better match for their skill-profiles (Jovanovic 1979) or to move onto steeper learning curves. In all these examples, economic actors choose new activities in non-random ways because these new activities are often related to the actor’s previous activities. The question of which industries are related and which are not is therefore essential to our understanding of a broad spectrum of economic phenomena.

Neffke and Henning (forthcoming) put forward a methodology that extracts information on the skill-relatedness among industries from labor flows. The rationale for the use of labor flows is that individuals are more likely to move among industries that are related than among unrelated industries. The authors’ skill-relatedness index proofs to be strongly predictive of corporate diversification patterns.

In this paper, we extend the work by Neffke and Henning by studying labor flows in separate segments of the labor market. Neffke and Henning propose that the differences in skill-relatedness among different groups of workers could yield interesting insights. However, the limited size of the Swedish labor market from which Neffke and Henning take their data precludes such an analysis. To overcome this difficulty, we use a linked employer-employee database that provides information on establishments and employees for the German labor market, the largest national labor market in the European Union.

We segment this labor market in four different ways. First, we compare the flows that can be considered to be local to those that would typically require a person to relocate to a new area. Next we compare the highest and the lowest paid segment of the labor force in each industry. A third segmentation is based on occupational groupings, distinguishing among management, sales, IT and other jobs. Fourth, we compare the flows in the labor force in East-Germany to the ones in the West-
German labor force. We use these segmentations to investigate to the extent to which the different labor market segments are characterized by different skill-relatedness networks.

Although we expected significant differences among the various skill-relatedness type, after correcting for measurement error, we find that, by and large, the different skill-relatedness networks are structurally very similar. What is more, when we confront our skill-relatedness estimates to Neffke and Henning's (forthcoming) dataset of diversifying Swedish firms, we find that they predict the direction of these firms’ diversification remarkably well, in some cases even better than Swedish relatedness indices. Taken together, these findings suggest that skill-relatedness is of a more general nature than we would have deemed possible.

In section 2, we discuss the concept and measurement of skill-relatedness and the rationale for the various labor market segmentations we use. Section 3 contains a description of the data. Section 4 describes and compares the skill-relatedness networks associated with each of the labor market segments and section 5 concludes.

2 Definition and measurement of relatedness

2.1 A resource-based definition of relatedness

The relatedness among economic activities has been studied most extensively in the strategic management literature (Penrose 1959, Teece 1982, Hoskisson and Hitt 1990, Montgomery 1994, Palich et al. 2000). These studies are often motivated by the question of how firms should diversify. In particular, firms can diversify into activities that are related or unrelated to their core activities. According to the resource based view of the firm (RBV: Wernerfelt 1984, Barney 1991) related diversification has some important advantages over unrelated diversification. An important antecedent to the RBV literature is Penrose (1959). She argues that firms possess resources from which they derive services that the firms use in their productive processes. However, the same resource can often provide different services, and the amount and quality of these services grow as the firm gains more knowledge about them through learning-by-doing. As a consequence, there will always be some services that are left unused. If these idle services can be used in other production processes than the firm’s current ones, the firm can grow organically by diversifying into new fields of activities that can absorb the idle resources. If we call these new fields related to the firm’s old activities, the relatedness among industries can be defined as the degree to which different industries use similar resources.

The above definition raises the question of which resources should be taken into consideration when measuring relatedness. There exist many types of resources, such as capital goods, financial resources, natural resources, technologies, human resources, etc.. Because, each of these resources can, in principle, be shared among certain sets of industries, Neffke and Henning (forthcoming) argue that there
may be as many types of relatedness as there are types of resources. It is therefore not surprising to find a wide range of different inter-industry relatedness measures.

Neffke and Henning (forthcoming) differentiate among three types of relatedness measures. The first type measures relatedness as the distance between the industry codes within the hierarchical structure of the industry classification system. For instance, two industries that are classified in the same narrow sub-sector are more related than two industries that only belong to the same broad overarching sector. We call such measures classification-based. The second type of measures, co-occurrence or outcome-based relatedness measures, assumes that the composition of industrial portfolios of productive units (e.g., firms, plants, or even countries or regions) reflect economies of scope. Therefore, if two industries are often found together in the portfolios of the same productive units, these industries are likely to be related. The third type, resource- or input-based indicators, directly investigates the similarities of industries in terms of their resource use. Important examples are indicators based on input-output matrices (Fan and Lang, 2000) and indicators that use information derived from patent data (Breschi et al., 2003). The skill-relatedness index of Neffke and Henning (forthcoming) we use in this paper is also of this type. It measures the degree to which human capital (i.e., human resources) can be shared among industries.

Neffke and Henning provide two reasons for focusing on human resources when assessing inter-industry relatedness. First, human capital is the prime resource for most firms in the modern economy. This view is supported by the knowledge-based view of the firm (KBV: Grant, 1996; Grant and Spender 1996), but it is also a widely held belief in the popular discourse about countries’ competitiveness. The second reason for developing a human capital based relatedness measure is that human capital is a very general resource that is used in all sectors of the economy. Therefore, the coverage of relatedness measures that reflect similarities in human capital requirements can be economy-wide. Other resource-based indicators are often biased to certain types of industries. For instance those measures that build on patent data can only measure relatedness among high tech industries, and input-output patterns are most useful to study relatedness among manufacturing activities. For an extensive discussion of the advantages and disadvantages of the different types of relatedness measures, we refer to Neffke and Henning (forthcoming).

2.2 Skill-relatedness

Skill-relatedness is based on cross-industry labor flows. An important justification for the use of labor flows is provided by the growing body of research in labor economics that show that human capital, i.e., skills and work experience, is specific to a person’s job. This specificity makes job switches costly. Scholars have shown that human capital is specific in various respects. The early work by Becker

---

1 For instance, policy discussions on regional and national competitiveness often give high priority to schooling and research and development, and these topics are also at the heart of the European Union’s Lisbon agenda.
highlights that human capital is firm-specific (Becker, 1962). More recently, Neal (1995) and Parent (2000) have provided evidence for the existence of industry-specific aspects of human capital and Poletaev and Robinson (1996), Gathman and Schoenberg (2010) and Nedelkoska and Neffke (2010) show that human capital is also strongly specific to the tasks that are carried out. As a consequence, individuals who change jobs risk rendering part of their human capital redundant. Moving to another firm would lead to the loss of firm-specific human capital. If the new firm also operates in a different industry than the old firm, this loss would be compounded by a loss of industry-specific human capital. In order to limit this kind of human capital destruction, individuals will switch to jobs in industries that value the same kind of skills as their previous employer. In other words, individuals will predominantly switch jobs between industries that are skill-related. For this reason, the size of the labor flow between two industries is a good indication of the extent to which the industries are skill-related.

Labor flows, however, also depend on a number of factors characterizing the industries other than their relatedness, such as the industries’ sizes, their growth rates and wage levels. Therefore, in order to judge whether or not a labor flow is exceptionally large, we need a baseline that takes the above mentioned industry characteristics into consideration. To this end, Neffke and Henning (forthcoming) construct a dataset with all possible pairs of industries and run a regression analysis of the observed cross-industry labor flows on a set of industry characteristics. The parameter estimates of this regression analysis are used to generate predicted flows, which serve as a baseline against which one can compare the observed labor flows. Following Neffke and Henning (forthcoming), we use a zero-inflated negative binomial model for these predictions. This model is well-suited to analyze labor flows, which constitute a count data phenomenon with an overabundance of zeros. Based on the predictions we obtain, we define the estimated skill-relatedness from industry $i$ to industry $j$, $\hat{S}R_{ij}$, as the ratio of observed to fitted flows:

\[(1) \quad \hat{S}R_{ij} = \frac{F_{ij}}{\hat{F}_{ij}}\]

where

$F_{ij}$: observed labor flow from industry $i$ to industry $j$

$\hat{F}_{ij}$: predicted labor flow from industry $i$ to industry $j$

Values over one indicate that industries are skill-related, whereas values between zero and one indicate that industries are dissimilar in their skill-use. Typically, the distribution of $\hat{S}R_{ij}$ is strongly right-skewed. Therefore, we will often use the following transformation, which maps the skill-relatedness measure onto the interval $[-1,1]$:


2.3 Differentiation of skill-relatedness by labor market segment

As noted above, Neffke and Henning (forthcoming) propose that relatedness may be a multifaceted concept, with different relatedness types reflecting the range of alternative applications of particular resources. In line with this, the authors suggest calculating different types of skill-relatedness that reflect the versatility of the human capital that is found in separate labor market segments. That is, industries that are connected by skill-similarities in one segment of their workforces may not be in other segments. In this paper, we use four ways to divide the German labor market into different segments in order to derive various types of skill-relatedness.

The first segmentation is motivated by Neffke and Henning’s concern that skill-relatedness may be confounded by the geographic distribution of industries. Most people try to find jobs close by home. As a consequence, industries that are often collocated in the same regions may display exceptionally large labor flows among them, not because the industries are skill-related, but because people prefer to move to nearby jobs. Therefore, we cannot rule out that our skill-relatedness matrices might mainly reflect the extent to which certain sets of industries locate in the same regions. In this case, our skill-relatedness measure would be flawed.

To investigate to what extent geography drives our relatedness estimates, we run a separate analysis for labor flows that would require a person to move to another region. We postulate that jobs that are over 100 km away from the old job typically require a person to relocate. This threshold reflects the fact that only 5% of German employees commute over distances greater than 50 km (Winkelmann 2010).

There are two reasons why we do not believe that the geographic distribution of industries will substantially distort our skill-relatedness estimates. First, our prior experience with the Swedish skill-relatedness matrices strongly suggests that the vast majority of skill-relatedness links can be attributed to similarities in skills. We expect to find similarly plausible skill connections in Germany. Second, even though job switchers may be limited in their choice to industries that offer local employment opportunities, individuals can still choose from a very wide range of industries in most German regions. Südekum (2006) reports a widespread structural convergence for German regions in the period 1993 and 2001. As a result, the majority of German regions have now approached the industrial structure that characterizes Germany as a whole. In addition, Dauth et al. (2010) report a long-term process of deconcentration of economic activities from 1975 to 2008 within West Germany. Nowadays it would seem that most regional economies in Germany consist of diversified sets of industries from which individuals can choose the ones that make the best use of their skills. As a consequence, we expect any differences
between skill-relatedness based on local and skill-relatedness based on non-local flows to be marginal.

The second segmentation is inspired by our earlier discussion on human capital specificities in labor economics and how these specificities lead to skill redundancies when people switch jobs. The extent to which people are exposed to such redundancies depends on the value of their skills: the more valuable they are, the more costly is leaving acquired skills idle. Wages provide a rough indication of the value of a worker's skills to her firm. Therefore, we split the labor force into in a group of employees that earn wages below the industry's median wage level and a group that is paid salaries above this level. On the one hand, low wage earners have less to lose in terms of human capital redundancies when switching to other industries. On the other hand, the skills of high wage earners, who typically are responsible for higher level and organizational tasks in a firm, may be more abstract and versatile than those of low wage earners. Such skills can probably be redeployed more easily across wide ranges of industries. The effects of the differences in level and generality may, in fact cancel out. However, even though low and high wage earners may be similarly flexible, the structure of the labor flows, i.e., the exact combinations of industries with high labor flows among them, is still likely to be very different. There is no reason to believe that the higher level skills of better paid workers will be transferable among exactly the same sets of industries as the lower level skills of worse paid workers. Therefore, we expect considerable differences between the high and low wage skill-relatedness matrices.

The third labor market segmentation aims at measuring relatedness type associated with different tasks in an organization. Penrose (1959) distinguishes among three major resource types that a firm can try to leverage when it diversifies, which correspond to three different tasks in a firm. The first are the entrepreneurial resources embedded in its management and used for organizational tasks. The second is the selling position the firm has acquired, which is needed to fulfill the firm’s sales task. The third resource type consists of the technological resources that make up a firm’s technology base and that are needed in the firm’s operational tasks. Following Neffke and Henning (forthcoming), we propose that the similarity among the skills used in different industries will depend on the exact task that an individual carries out. Management and sales personnel, for instance, might possess skills that are more transferable across industries than individuals that work in a firm’s production processes. Furthermore, it could be that industries that are skill-related in some tasks are not in others. For instance, although making cars may be similar to making trucks, selling cars may require different skills from selling trucks because both types of vehicles use different sales channels. Finally, we add a fourth task that is important in many firms, managing the information technologies (IT). The reason for adding IT tasks is that IT is an important general purpose technology that is found in almost every economic activity. Given the generality and distinctiveness, IT activities may exhibit different relatedness patterns from the other activities.
The skill-relatedness types corresponding to each task are estimated by splitting the labor force along the lines of occupational groupings. This management task is captured by occupations that are involved in higher or lower management type tasks, including for instance, foremen and naval officers. Because the sales task of a firm requires marketing skills, we include occupations related to advertising, such as public relations and design among the sales occupations. The IT task is identified by isolating flows of IT personnel. The remaining occupations are treated as a separate group and are supposed to capture the skills needed to run the daily operations in an industry. Table 1 contains a list of the selected occupations for each of these segments.

TABLE 1 ABOUT HERE

The role managers play in a firm is very different from the one of the general workforce. Although knowledge of the production technology is important, what is essential in the management task is the coordination of the efforts of individual workers. Similarly, affinity with a firm’s product will benefit sales efforts, but marketing and sales departments require different skills from the operational and management skills in a firm. Finally, even though certain IT challenges are industry-specific, it would seem that, in most industries, many valuable skills of IT specialists are not strongly industry specific. Because of these considerations, we believe that this occupational segmentation is likely to yield radically different relatedness types.

The final segmentation reflects Neffke and Henning’s suggestion that the relatedness among industries may depend on the cultural and institutional background against which skills are accumulated. Reunified Germany offers a remarkable test case, where the eastern part has historical roots in a radically different techno-economic institutional setting compared to the western part. The institutions of East Germany’s socialist past have left deep traces. After the reunification in 1990, institutional structures of West Germany were introduced stepwise in the eastern regions. The transformation of East Germany’s economy was shaped by the implementation of western political, social and educational systems as well as by the adoption of West Germany’s labor market administration and institutions. Even though the reunification took place some fifteen years before the period we study, the old institutions of the former GDR still play a role because a considerable part of the labor force received its education and built up work related skills in the period before 1989. The effect of East Germany’s institutions may have endured in as far as skills have been transferred to younger employees and in as far as the institutions have shaped more persistent cultural idiosyncrasies. Be this how it may, the fact that many of the regional economies of the East still have a long way to go before they reach the average prosperity levels that western regions enjoy shows that the economies of the east and the west are very different even today. A large body of research collected in a recent special Issue of *Informationen zur Raumentwicklung* (2010) and in Krause et al. (2010) confirms the existence of these
long-lasting, yet converging structural differences, highlighting the political, demographic and socio-economic disparities between eastern and western regions twenty years after the German reunification. If skill-relatedness is not so much the result of the state of production technology, but is to an important extent a social construct, the slow convergence of institutions may translate into different skill-relatedness structures between East and West Germany. Even though we are sympathetic toward this line of reasoning, we believe that the productive processes in East and West Germany should be more or less similar in their organization and skill requirements. Therefore, although we would expect moderate differences, we do not believe these differences to be larger than the ones between high and low wage earners.

To summarize, we use four different ways to divide the German labor market into different segments and then use these segments to construct estimates for various skill-relatedness types. We expect the largest differences in the segmentation by occupational groupings. We also think that the high wage-low wage distinction may result in two quite disparate skill-relatedness types. Although we are less convinced, we might also see differences between the East German and West German skill-relatedness estimates. We expect the smallest differences in the skill-relatedness based on local and non-local flows.

TABLE 2 ABOUT HERE

In all the different labor market segmentations described above, the variables in the regression equation that is used to calculate the baseline flows relate to the labor market segment of an individual's old job. That is, a labor flow is defined by the segment from which the flow originates, not by the one in which it ends. In the segmentation into the labor forces of East and West Germany, we exclude flows between the two regions. We believe that the migration between East and West Germany deserve a study in its own right. Furthermore, we require that an industry employs on average at least 500 employees per year in a labor market segment. Smaller industries are left out of the analyses, because they exhibit flows that are too small for a meaningful assessment of relatedness. The various segmentations give rise to labor market segments of different sizes. Table 2 displays the number of employees in each labor market segment for the years 2003 to 2006. The differences among the totals of different segmentations are due to a small number of cases where the information on individuals' occupations and geographic locations is missing.
3 Data

3.1 Description of the Database

We derive cross-industry labor flows from the Historic Employment and Establishment Statistics (HES).\(^2\) This database relies upon German social security records. Employers are obligated by law to provide information for all employees who are employed during a year at the level of individual establishments. The employers report this information to the social insurances, which inform the Federal Agency of Employment on four yearly due days at the end of each quarter (March 31, June 30, September 30 and December 31) about the daily wage on the respective due day and about a range of socio-demographic variables (such as the attained educational level, nationality, gender, and age). The wage information is very reliable, because this information determines the social security contributions. However, the wage information is censored due to a contribution limit to the social security. Additionally, information is available on industry, occupation and work status (full-time, part-time, apprenticeship), places of work and residence. Individuals and their employers can both be followed through time with the help of anonymized identifiers. These identifiers link individual and establishment data into employer-employee datasets. This database covers the years from 1975 to 2010, allowing the observation of long-term work histories of individuals and developments of establishments over a maximum span of 36 years. Crucial to this study, job switches of individuals between establishments, industries and regions can be tracked over this entire period.

The industry classification system of this database has seen three major overhauls since 1975: one in 1993, another in 2003 and the final one in 2008 (see Table 3). The Statistical Classification of Economic Activities in Germany 1973 (WZ 1973) is valid from 1975 to 1998. This industry classification is constructed using a nested hierarchy of three different levels. The later classification systems, WZ 1993, WZ 2003 and WZ 2008, are built up in a hierarchy of six levels. Each industry has a 5-digit code. Based on the first two digits of these codes, industries are grouped into sets that are coded by two capital letters. The first letter denotes the section to which the industry belongs and the second letter provides a further division in sub-sections.\(^3\) The remaining four levels are indicated by the first 3, 4 and 5 digits of the code. In this article, we use the terms "sectors" (sets of industries in the same section), "sub-sectors" (same sub-section) 3-, 4- and 5-digit industries (sets of industries for which the first 3 or 4, respectively all 5 digits are the same). All classification systems are based on a nested hierarchical structure. As a

\(^{2}\) We refer the reader to Bender, Haas and Klose (2000) for a detailed description of this database.

\(^{3}\) In the WZ 2003 classification for instance, the industry “Manufacture of parts and accessories for motorcycles” is coded as industry 35412. This industry is a part of sub-section DM (“Manufacture of transport equipment,” which in turn belongs to section D (“Manufacturing”). In the finer grained classification, the industry is part of the 2-digit industry 35 “Manufacture of other transport equipment”, the 3-digit industry 354 “Manufacture of motorcycles and bicycles” and the 4-digit industry “Manufacture of motorcycles.”
consequence, lower classification levels can always be aggregated to higher ones. The WZ 1993, WZ 2003 and WZ 2008 are harmonized with the European NACE (Nomenclature générale des Activités économiques dans les Communautés Européennes) classification at the 4-digit level. Therefore, most of our calculations do not use finer-grained classification levels than the 4-digit codes. Occupational information is reported in the Historic Establishment and Employment Database in accordance to the Occupational Classification of the Federal Agency of Employment in Germany (Bundesagentur für Arbeit, BA). At the lowest level of aggregation, about 340 occupations are identified by 3-digit codes.

Hethey/Schmieder (2011) state that the HES is more comprehensive and more accurate than comparable datasets such as the Unemployment Insurance Data or the LEHD in the US. However, the database has some limitations, due to the fact that it covers only those employees that are subject to obligatory social insurance in any establishment in Germany that employs at least one employee. Therefore, there is no information on individuals who are not subject to social insurance contributions such as civil servants, soldiers, self-employed individuals, entrepreneurs and unpaid family workers. Yet, the database still covers 80 percent of the total employment stock in Germany (Herberger and Becker, 1983).

At the level of establishments, these limitations mean that enterprises with a workforce that consist only of the owner-manager are not included. In addition, the coverage of the agricultural industry in the HES is weak, because the industry’s workforce mostly consists of self-employed persons. Further limitations arise due to the regularities and routines of assigning establishment identifiers by the Federal Employment Agency. These issues are explained in detail in Hethey and Schmieder (2010) and in Fritsch and Brixy (2004).

For our purposes, we need to identify the cross-industry labor flows that occur on the German labor market. We define the labor force as all employees with a full-time job between the ages of 18 and 65. Furthermore, we do not consider employees with a part-time job, unpaid family workers, apprentices, trainees and interns. The original data are stored at the level of employment spells, yielding several rows of data per individual per year. To reduce the computational burden, we retain only the information on an individual’s employment situation once a year, on 30 June. This due day is less affected by seasonal effects than the alternatives.

3.2 Labour Flows

We define job switches as events in which an employee works in one establishment on 30 June of one year but in another establishment on 30 June of the next year. The labor flow from an industry $i$ to an industry $j$ is calculated as the sum total of employees that change jobs from establishments in industry $i$ to establishments in industry $j$. This definition requires that an employee not only changes industries, but also establishments. This way, we are certain that the employees of establishments that change industry affiliations are not mistakenly included in the inter-industry
labor flows. Unfortunately, however, the establishment identifiers in the data set are not completely reliable. In a study of exit and entry rates in Germany, Hethey and Schmieder (2010) find that only 35% to 40% of all establishments with more than three employees that had new or disappearing establishment identifiers can be unambiguously labelled as entries and exits, respectively. Moreover, establishments that are spun out off existing establishments may yield large labor flows from the old establishments’ industries to the spin-offs’ industries. However, one can argue that these labor flows represent much weaker indications of skill-relatedness than labor flows that result from individuals’ own, private, job change decisions. To confront these problems, we analyze the chunkiness of flows between establishments in an effort to determine which labor flows may be spurious. This strategy resembles Hethey and Schmieder’s (2010) approach. The underlying rationale behind these analyses is that the more people move en block from one establishment to another the more likely it is that these moves result from organizational restructuring processes, such as spin-offs, mergers or break ups of business units or simply recordings of establishments and not from individuals’ conscious decisions to move from one establishment to another. Our main interest is identifying cross-industry labor flows and not accurately assessing the stability of the establishment coding system. We therefore content ourselves with identifying those labor flows that are most likely to be spurious. These labor flows are subsequently dropped from our analyses. In other words, if large chunks of the labor force of one establishment either come from or move to the same other establishment we omit these individuals from the labor flows that we use for constructing relatedness estimates.

FIGURE 1 ABOUT HERE

Whether a flow should be considered as “chunky” depends on the size of the establishment in which the flow originates and of the one in which it ends. For establishments with fewer than five employees, a flow is considered chunky if all of the employees either came from or went to the same establishment. For plants with five or more employees, flows are considered to be chunky if they represent at least 80% of the employment in the establishment where the flow originates or in the establishment of the flow’s destination. Furthermore, flows of 100 employees or more are considered chunky regardless of establishment sizes. Figure 1 illustrates this definition.

Apart from flows that are spurious, also overhauls to the industry classification system present some challenges. As described in section 3, the classification system was renewed on three occasions in the period 1975 to 2008. Although correspondence schemes exist, we chose not to rely on these schemes. Instead, we extrapolate each classification system for as many establishments as possible. This is possible, because often years exist for an establishment in which both the old and the new industry codes are provided. Take for instance an establishment for which in 1993 both the WZ 1973 and the WZ 1993 codes are provided. If the establishment also existed in 1992, and the WZ 1973 code is the same as the one in
1993, but the WZ 93 code is missing, we copy the WZ 93 code from 1993 for the establishment in 1992. This allows both forward and backward filling of industry codes for as long as the establishment does not change the industry code in a year’s main classification system. The disadvantage of this method is that establishments that exit the data set before a new classification system becomes available will not be classifiable in other classification systems. Similarly, there is no way to provide a classification code for establishments that enter the data set after a classification system has been abandoned. Any establishment that changes industry affiliation will also be impossible to reclassify. The advantage is that skill-relatedness is not affected by imperfections in correspondence tables. This method allows us to identify the WZ 03 industry codes for 7.5 million employees as far back in time as 1975 and of 19.6 million employees as far forward in time as 2008. Even in 1975 we can work with a larger number of employees than Neffke and Henning (forthcoming) dispose of in their Swedish study. In this paper we focus on the construction of the skill-relatedness estimates for Germany. We therefore restrict our descriptive analyses to the WZ 03 classification and the period in which this classification has been in use (i.e., 2003-2007). In future work, we plan to exploit the entire time period to investigate the dynamic aspects of skill-relatedness. To enhance comparability across countries, our main focus lies with the 4-digit level, which is harmonized for member states of the European Union in the NACE Rev 1.1 classification.

3.3 Cross-industry labour flows

Table 3 summarizes the degree to which labor flows cross industry boundaries at different levels of aggregation. The table refers to the situation in Germany as a whole. Cross-industry labor flows are very common. Moreover, 60.4% of all employees who change jobs between two different industries, do not just move to different 5-digit industries, but to entirely different sectors. To some extent, this may simply reflect that the possible combinations of two 5-digit industries that belong to different sectors are far more numerous, than the possible combinations that consist of 5-digit industries in the same sector. In fact, 81.3% of all possible 5-digit industry pairs do not belong to the same sector (first column of Table 3). The second column describes the switching behavior that would be expected if individuals chose new industries randomly with the probability of choosing a particular destination industry proportional to the industry’s employment size. The random benchmark for flows among industries in different sectors is 85.7%. This is a much higher proportion than the 60.4% we observe, showing that people do not move randomly among industries. This non-randomness is best illustrated for the flows among industries that belong to the same 4-digit industry class. The flows in this category are about 25 times higher than predicted by the random baseline. It is also interesting to note that the high-wage earners display a movement pattern that is even more distinct from the random pattern than the overall population of employees. Of all high

4 For this year, we only have West-German data.
income earners that switch industries, 8.1% remain in the same 4-digit industry class. At 6.3%, the corresponding proportion for low wage earners is substantially lower. This is consistent with Neffke and Henning’s (forthcoming) contention that high wage earners are more likely to protect their human capital investments when changing jobs. However, against the random baseline of 0.3%, also the low wage earners’ job switching behavior is far from random.

Table 4 provides similar information for various labor market segments. Because of the reduced number of employees on which we now base relatedness, we drop the 5-digit level of the classification system. The flows in East and West Germany are comparable in the sense that the extent to which the employees in both regions cross sector borders is virtually indistinguishable. However, in the eastern part of the country, people are much more likely to stay in the same 2-digit and even 3-digit classes than in the western part. This holds for high and low earners alike.

The figures in the third through sixth column of Table 4 indicate that large differences exist among the labor flows of different occupational groupings. Switching among sectors is least prevalent in sales occupations. Indeed, these occupations are associated with infrequent cross-industry switches. The other category that stands out is IT. Although IT workers show average switching behavior across sectors, the employees in this segment tend to remain in the same subsector more than the employees in the other segments do. Finally, differences between the patterns of local and non-local labor flows are negligible. In the next section, we investigate whether these similarities and dissimilarities in the labor flows of different labor market segments can be attributed to differences or the lack thereof in different skill-relatedness types.

4  Skill-relatedness
4.1  Industry space
We calculate skill-relatedness for the various labor market segments using the method explained in section 2.\textsuperscript{5} We start by analyzing the skill-relatedness of the German labor market as a whole. Figure 2 presents the findings in the form of a

\textsuperscript{5} Due to space restrictions, we will not discuss the outcomes of the regression models that are used to predict the baseline labor flows across industries. We naturally find that larger industries experience both larger in- and outflows of labor from other industries. A strong employment growth in general depresses labor outflows but raises the inflows of labor from other industries. High wage levels are in most regressions associated with a lower labor flows in either direction. As expected however, in particular the outflow of labor is lower for higher income industries. Moreover, regressions in which higher wages are associated with higher outflows, the concomitant increase in inflows is even stronger. The results of these estimations are available upon request.
network of industries, which we will refer to as the industry space. The nodes represent 4-digit industries and the colors and shapes of the nodes correspond to the sectors to which industries belong. The size of the nodes depends on the size of the industries' workforces. The presence of a link between two industries indicates that the industries are strongly skill-related. In principle, industries are skill-related if their SR value exceeds one. However, depicting all such links would clutter the image. Therefore, we plot only the strongest links. We follow the rule of thumb suggested in Hausmann et al. (2011) that an average of four links per industry gives good visual results. To avoid isolated nodes, we also make sure that each industry is connected to at least one other industry. The positions of the nodes are determined by a spring-embedding algorithm, which groups sets of closely interlinked nodes together in the two-dimensional plane.

FIGURE 2 ABOUT HERE

Figure 2 shows that industries that belong to the same sector are often skill-related. For instance, the industries in the financial sector (yellow squares) all cluster in the upper right corner. The purple squares, representing hospitality industries, are all found in the lower right corner of the network. However, this clustering of industries by sector is not absolute. There are many connections between industries that belong to different sectors. The manufacturing industry that produces medical equipment (center left) is connected to the dentists in the care sector and to the wholesale and retail industries involved in the distribution of medical products. Similarly, the construction industry of plumbing (lower right) links to the manufacturing industry that produces cooling equipment. Moreover, the industry space shows the relatedness among entire sectors. Table 5 gives an overview of the top 20 strongest related pairs of industries.

TABLE 5 ABOUT HERE

Our analyses corroborate the finding in Neffke and Henning (forthcoming) that “skill-relatedness links industries that intuitively would seem to use similar skills, even though traditional industry classification systems place these industries into

---

6 The labels in the graph are abbreviated industry names. A list of full names and corresponding industry codes are available from the authors.

7 The exact procedure is as follows. We first eliminate all links that are insignificant at the 5% level. Next, we calculate a cut-off value at which there are, on average, four links per node and keep the links that exceed this threshold. In the final step, we add the link that connects any isolated industry to its most skill-related neighbor.

8 We use the NetDraw (Borgatti, 2002) software programme to construct network graphs.

9 One will notice that skill-relatedness is asymmetric, yielding different values for the relatedness between industry $i$ and industry $j$ and the one between industry $j$ and industry $i$. Although this asymmetry may have important implications about the generality and complexity of the skills associated with an industry, we leave this topic for future research.
profoundly different categories.” Before turning to the comparison of the indices of different skill-relatedness types, we must first tackle a problem that arises due to measurement error.

4.2 Dealing with noisy skill-relatedness estimates

Neffke and Henning (forthcoming) already concede that skill-relatedness measurements are subject to substantial measurement error. In particular among small, industries, the baseline flow in the denominator of equation (1) is very small, sometimes only a fraction of one. As a consequence, a small increase in the numerator, which consists of the observed labor flows between two industries, will lead to very large changes in skill-relatedness. One solution we adopt to deal with this noisiness is aggregating industries to the 2-digit level. However, this solution comes at the cost of aggregating industries with possibly heterogeneous skill-requirements, compounding any aggregation biases. Another solution is dealing with the measurement error more directly, by estimating the noisiness of our measures and correcting the correlation for the associated attenuation bias. We present two different procedures for this noise correction.

Let us assume that the observed skill-relatedness between two industries in a given year consists of two parts: an unobserved real part and an identically and independently distributed measurement error:

\[
\overline{SR}_{ij} = SR_{ij} + \epsilon_{ij}
\]

In this expression, \(\overline{SR}_{ij}\) is the transformed estimated skill-relatedness of industry \(i\) and \(j\), whereas \(SR_{ij}\) is their real transformed skill-relatedness and \(\epsilon_{ij}\) the measurement error.

For ease of notation, let us use the lowercase letters \(x\) and \(y\) for the vectors that are obtained by stacking the columns of the transformed matrices for two different types of skill-relatedness, while omitting the diagonal elements:

\[
x = \begin{pmatrix}
SR_{12}^{\text{type1}} \\
\vdots \\
SR_{1N}^{\text{type1}} \\
\vdots \\
SR_{N1}^{\text{type1}} \\
\vdots \\
SR_{N(N-1)}^{\text{type1}}
\end{pmatrix} \quad ; \quad y = \begin{pmatrix}
SR_{12}^{\text{type2}} \\
\vdots \\
SR_{1N}^{\text{type2}} \\
\vdots \\
SR_{N1}^{\text{type2}} \\
\vdots \\
SR_{N(N-1)}^{\text{type2}}
\end{pmatrix}
\]
One can compare the relatedness types 1 and 2 by calculating the correlation between the observed values of $x$ and $y$ in year $t$, $\text{Corr}[\hat{x}_t, \hat{y}_t]$. Using equation (3), we can write this correlation as follows:

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

In this expression, $\varepsilon_{xt}$ and $\varepsilon_{yt}$ denote the measurement errors associated with the skill-relatedness of types 1 and 2 respectively in year $t$. Assuming that these measurements are independent from one another and from the real skill-relatedness vectors $x$ and $y$, we can rewrite this expression to arrive at the following equation (a detailed derivation is provided in Appendix 1):

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

The first right-hand side term is the correlation between the real skill-relatedness of type 1 and 2 we set out to find. Therefore, if we can somehow quantify the relative size of the measurement errors, it should be possible to quantify the real association between the skill-relatedness types. For this purpose, we would need two independent measurements of the same skill-relatedness type. If we are willing to assume that measurement errors are not only independent from each other and from the real skill-relatedness values, but that they are also uncorrelated through time, we can repeat the skill-relatedness estimates for each year and use these as independent measurements. Assuming that the variance of the measurement error does not change substantially between years $t$ and $t'$, $\varepsilon_{xt} = \varepsilon_{xt'} = \varepsilon_x$ holds. Now, it is straightforward to show that:

\[
\text{Corr}[\hat{x}_t, \hat{x}_{t'}] = \frac{\text{Var}[x]}{\text{Var}[x]+\text{Var}[\varepsilon_{xt}]}
\]

Denoting the correlation between variables $a$ and $b$ by $\rho_{ab}$, (5) simplifies to:

\[
\rho_{xy} = \frac{\rho_{x't'}\rho_{yt'}}{\sqrt{\rho_{x't'}\rho_{x't'}}\sqrt{\rho_{yt'}\rho_{yt'}}}
\]

Given that we find that differences between all pairwise correlations of the four different measurements we can obtain for each skill-relatedness type, this assumption seems harmless. The full derivation can be found in Appendix 1.
This correction is well-known and was first proposed by Spearman (1904). The assumptions that the error terms are uncorrelated with each other and through time may be too stringent. Another way to reduce the impact of measurement error is by averaging the relatedness measures over a suitably large number of years. Unfortunately, we will need many more years of observations than the four we have. Moreover, this strategy also presupposes that skill-relatedness is unchanging and forever given. We do not think this is likely to be true. However, if we are willing to assume that relatedness does not change much in our four year time window, we can also first average out part of the measurement error and only then control for the remaining error. In Appendix 2, we show that the following alternative correction can be achieved with our data (this correction is similar to another proposal of Spearman (Spearman, 1910)):

\[
\rho_{xy} = \rho_{\bar{x}\bar{y}} \sqrt{1 + \frac{1}{t} \left(1 - \rho_{\bar{x}_{h}^{2}}\right) \left(1 - \rho_{\bar{y}_{h}^{2}}\right)}
\]

In equation (9), \(T\) denotes the number of different years for which relatedness is available and \(\rho_{\bar{x}\bar{y}}\) the correlation between the averages of the two types of skill-relatedness over these years. As long as none of the assumptions on the measurement errors are violated, equations (8) and (9) should give the same results. Similar findings for both methods should thus reduce concerns about the assumptions being violated.
4.3 Comparing skill-relatedness types

We are now in a position to assess the similarity of the various skill-relatedness types. Before looking at the noise-corrected correlation coefficients among the various skill-relatedness types, we inspect their industry spaces visually. Figure 2 shows that such a visual analysis will be unwieldy at the 4-digit level. Therefore, we plot all industry spaces at the 2-digit level. The comparison between high and low wage segments is shown in Figures 3a and 3b. The West and East German industry spaces are depicted in Figures 4a and 4b. The industry spaces related to management, sales, IT and other tasks are shown in Figures 5a to 5d. Finally, Figures 6a and 6b contain the industry spaces based on local and non-local flows respectively.

At first sight, the industry spaces look rather dissimilar. However, this impression is misleading. Even in rather similar networks, the visualisation algorithm may assign very differently coordinates to the nodes. However, for similar networks, nodes’ relative positions vis-à-vis each other will be, by and large, the same. That is, the similarity of two industry spaces will reveal itself in the extent to which the same sets of industries cluster together. By this criterion, the similarity of all of the industry spaces is remarkable. For instance, not only do the yellow squares of the industries in the financial sector cluster together in every single graph, they are also located relatively close to the orange squares of the business services in each industry space. The green squares of the transport industries form similarly tightly knit clusters in all graphs and distinct groups of manufacturing industries (such as the furniture, apparel, leather and textiles industries) are connected in every single graph.

This first visual inspection suggests that, counter to our expectations, the different relatedness types yield rather similar industry spaces. This happens even though industries differ massively in employment in different labor market segments, as evident in the size differences of the nodes in different industry spaces. For instance, sales jobs are predominantly found in only two 2-digit industries – wholesale and retail – and the IT industry space is dominated by the otherwise tiny industry of computer related business services. Against this background, the finding that skill-relatedness always connects the same industry pairs is remarkable. In order to quantify this similarity, we turn to the correlations between the different skill-relatedness types. For this purpose, we calculate the matrices that contain transformed skill-relatedness types for each year in our dataset separately, stack their columns into long vectors and drop the elements that correspond to the
matrices’ diagonals. Tables 6a and 6b report the raw and the attenuation-corrected correlations at the 4- and 2-digit level respectively.\textsuperscript{12}

**TABLES 6a and 6b ABOUT HERE**

The upper panels of Tables 6a and 6b present the correlations for the skill-relatedness estimates of successive years at the 4- and 2-digit level, respectively. These estimates fall far below one, suggesting substantial measurement errors. In line with this finding, the raw cross-type correlation coefficients are moderate, ranging from 0.30 to 0.50 for 4-digit industries and from 0.45 to 0.75 for 2-digit industries. However, if we correct for the noisiness of the skill-relatedness, we find that cross-type correlations are remarkably high. The lower panel summarizes the correlation coefficients that result with the noise correction as described in (8). The uncorrected cross-type correlations between four-year averages are higher than their one-year counterparts (typically between 0.55 and 0.70 for 4-digit industries and between 0.65 and 0.85 for 2-digit industries), confirming a strong influence of measurement error. In parentheses, we provide the correlation coefficients using the attenuation correction based on these averages as proposed in (9).

Both correction methods yield outcomes that lie remarkably close together, suggesting that the underlying assumptions are not violated. The cross-type correlations now range from 0.74 to 0.97 for 4-digit industries and from 0.80 to 0.97 for 2-digit industries. In fact, the outcomes at the 2-digit level are very similar to the ones we find at the 4-digit level.\textsuperscript{13} This confirms our earlier observation that the industry spaces of different relatedness types are remarkably similar. The difference between the high and low wage segments is negligible, both for Germany as a whole and for the East and West Germany separately. Local and non-local flows yield slightly different skill-relatedness matrices, but at 0.89, these differences are minor. In fact, even the different matrices yielded by the occupational segmentation display correlation coefficients that are beyond what we imagined possible. For instance, with a correlation of somewhere between 0.87 and 0.94, the skill-relatedness for IT and management is virtually the same. Apparently, regardless of whether workers are responsible for the management, sales or IT support in an industry, they are likely to move to jobs in the same set of other industries. Counter to our expectations, the lowest correlations we find are between the East and West German skill-relatedness matrices. This suggests that skill-relatedness is more affected by institutions than by the tasks that were carried out at a firm. However, given the high correlation levels, we are rather led to the surprising conclusion that skill-relatedness displays a high level of universality. To conclude our empirical

\textsuperscript{12} In order for the bias correction to work properly, we make sure that the samples used to calculate correlations between two consecutive measurements of the same relatedness type and the ones used to calculate cross-type correlations are every time the same, while always retaining the largest available number of industry combinations.

\textsuperscript{13} In fact, the outcomes at the 2-digit level are very similar to the ones we find at the 4-digit level, which further validates the bias correction methods we used.
analyses, we probe deeper into this universality thesis by subjecting the skill-relatedness measures to an external validity test.

4.4 External validity: predicting corporate diversification in Sweden

In section 2, we noted that the resource based view of the firm predicts that diversifying firms will choose diversification targets that are closely skill-related to their core activities in order to leverage their human resources. For a sample of 649 events in which Swedish firms built new plants in industries they were not yet active in, Neffke and Henning show that firm diversification is indeed often strongly skill-related. We use this Swedish sample to investigate to what extent the various German relatedness types also classify each of the 649 diversification as skill-related.

First, we construct one additional relatedness variable that captures the extent to which different industries are found in the same parts of the industry classification hierarchy as a benchmark. Accordingly, 4-digit industries are assumed to be completely unrelated if they are classified in different sectors and they are most related if they are part of the same 3-digit industry. In this way, the four hierarchical levels by which 4-digit industries are classified yield five different relatedness classes. Using the European designation of the classification system, we refer to this relatedness as NACE relatedness. The number of combinations of two industries that belong to different sectors is far larger than the number of combinations for which the industries belong to the same 3-digit industry (191,202 versus 1,446 combinations). Therefore, the NACE relatedness categories vary widely in size. In order to be able to compare the skill-relatedness types to the NACE relatedness, we turn the continuous skill-relatedness variables into categorical variables. Categories are chosen in such a way that each type contains the same share of all industry combinations for which the relatedness type could be calculated.14 These shares are shown in the first column of Table 7. The other columns of this table count the number of diversification that fall into each class for every skill-relatedness type.

| TABLE 7 ABOUT HERE |

The NACE relatedness column corroborates the finding of Neffke and Henning that the industry classification system offers a poor guide to the relatedness of diversification moves. Over two thirds of these moves take place among industries in different sectors of the economy, suggesting that firm diversification is largely unrelated. Based on skill-relatedness, however, firms seem to diversify mostly in

14 Given that we require an industry to employ at least 500 individuals per year, the number of industry combinations for which relatedness could be calculated differs from one labor market segment to the other.
industries that are highly related to their core activities. In fact, although category 5 contains only 0.6% of all industry combinations, for most skill-relatedness types, it hosts over 15% of all firm diversification events. Even though the first category, which contains the most unrelated industry combinations, covers over 125 times more industry pairs, we find only twice the number of diversification moves, around 30%, in this category. It is important to note that the diversification moves took place in Sweden, whereas the labor flows underlying the skill-relatedness estimates took place in Germany. The only plausible explanation for this finding is that firm diversification and labor flows are driven by the same universal structure that connects industries to one another and that we called skill-relatedness.

5 Conclusions, caveats and future research

5.1 Conclusions

We started out by the hypothesis that different sets of industries will share similar skill-requirements when focusing on different segments of the workforces of these industries. However, our evidence points in the opposite direction. Instead of finding different types of skill-relatedness, our evidence suggests that there is a single general skill-relatedness that affects the labor flows in all segments of the labor market. This finding is remarkable, because we found substantial differences in raw labor flows for the various labor market segments. However, once controlling for size, growth rate and wage effects on the one hand and measurement error on the other, our findings suggest that the overlaps of industry specific knowledge and skills among industries are experienced in virtually the same way throughout the entire workforce. We will now discuss these findings and point out some caveats and directions for future research.

5.2 Caveats: the role of social networks in job search, the low signal-to-noise ration of skill-relatedness, and the universality of skill-relatedness

Neffke and Henning (forthcoming) already note that social networks play an important role in finding new jobs (Granovetter, 1973; Lin and Dumin, 1986; Wegener, 1991; Fernandez et al., 2000). To the extent that our skill-relatedness networks reflect the strength of social networks among industries, this could explain why we find that different occupational groups have similar skill-relatedness networks. However, since social networks are often very local, we would expect that non-local flows, in which workers cannot draw as strongly on their networks, should yield disparate skill-relatedness types. However, our findings show that such differences are minor. Given that an important reason for industries to collocate is sharing a common labor pool, the non-local flows are actually biased against flows among skill-related industries. Therefore, it is not surprising to find some differences among the segments in this segmentation.
diversification patterns of Swedish firms. In fact, some of the measures even outperform the Swedish skill-relatedness measure.

Another important caveat is that the yearly skill-relatedness measures have a signal to noise ratio that is typically about $\frac{0.4}{0.6} = \frac{2}{3}$. Even though a general skill-relatedness structure may exist, there seems to be ample scope to deviate from this structure when choosing a new job. At this point, we cannot say whether such deviations represent inefficiencies in the labor market. To answer this question, one would need to investigate whether people that build careers of jobs in skill-unrelated industries perform worse than those that follow the skill-relatedness structure more closely. Indeed, because these careers would be more challenging and provide opportunities to combine skills that are otherwise often found in quite disparate industries, a select group of people that is able to successfully follow such a career path may even reap substantive benefits.

We believe that some of the labor market segments are different so substantially from one another that the finding of a general underlying skill-relatedness structure in each of them is surprising. However, we cannot exclude nor do we believe that skill-relatedness is absolutely universal. We expect that large differences will be found in economies that are at different stages of development. It would be highly interesting to construct an industry space for China or India, for instance. Moreover, given the changes in technology witnessed in the past 100 years, we also expect to see large shifts in the relatedness structures over time. These shifts may accompany large technological changes, as the shifts in techno-economic paradigms envisioned by Freeman and Perez (1989). To investigate such phenomena, one needs to draw on historic data and somehow overcome the considerable difficulties that arise when classification systems change. However, the rewards of these efforts are potentially large, as they may shed new light on how technology evolved (in particular so-called general purpose technologies) and how changes in relatedness structures may have favoured some economies over others.

5.3 Future research

In future work, we want to address a number of issues. Another avenue for future research is into the consequences of skill-relatedness. At the level of regional production portfolios, the notion of related diversification has recently gained traction in economic geography. Many of the original ideas of Penrose have been transferred to the setting of regional economies through the concept of regional branching in evolutionary economic geography (EEG). Similarly, Porter’s notion of clusters of related industries (Porter 2003) and Florida’s recently coined term of geographies of scope attribute an important role to the inter-relatedness among activities in a region in understanding a region’s economic fortunes. A second exciting opportunity is the above mentioned possibility to explore skill-relatedness’s

---

16 Neffke et al. (2011) demonstrate regional branching empirically by following a study of Hidalgo et al. (2007) about the diversification patterns of national export portfolios.
temporal dimension. Over the past four decades, the economy underwent some important changes. For instance, we expect that the general shift from manufacturing to services and the gradual expansion of IT will have left an imprint on the relatedness structure. Such analyses will help gain a more profound understanding of to what extent the skill-relatedness structure in the economy constrains and shapes the direction of economic development.
References


Neffke, Frank; Henning, Martin (2010): Seeds of regional structural change. The role of entrepreneurs and expanding firms in shaping local path dependencies, Papers in Evolutionary Economic Geography, 10.05, Utrecht University.


Appendix 1: Derivation of correction for attenuation bias based on a single cross-type correlation

In this section, we derive the bias correction given in equation (8). A similar correction was proposed first by Spearman (1904). Using (3), the correlation between two skill-relatedness types, $x$ and $y$, measured in year $t$ can be written as follows:

\begin{align}
\text{(A1)} \quad \text{Corr}[^x_t,^y_t] &= \frac{\text{Cov}[x+\epsilon_xt,y+\epsilon_yt]}{\sqrt{\text{Var}[x+\epsilon_xt]\text{Var}[y+\epsilon_yt]}} \\
\text{(A2)} \quad \text{Corr}[^x_t,^y_t] &= \frac{\text{Cov}[x,y]+\text{Cov}[x,\epsilon_xt]+\text{Cov}[\epsilon_xt,y]+\text{Cov}[\epsilon_xt,\epsilon_yt]}{\sqrt{\text{Var}[x]+2\text{Cov}[x,\epsilon_xt]+\text{Var}[\epsilon_xt]}(\text{Var}[y]+2\text{Cov}[y,\epsilon_yt]+\text{Var}[\epsilon_yt])}
\end{align}

Let us now assume that the measurement errors are uncorrelated with the real skill-relatedness values of both types:

Assumption 1a: \(\text{Corr}(x, \epsilon_xt) = \text{Corr}(y, \epsilon_yt) = 0\)

Assumption 1b: \(\text{Corr}(x, \epsilon_yt) = \text{Corr}(y, \epsilon_xt) = 0\)

Because skills are latent constructs in our treatment, Assumption 1a can be regarded as a definition: whatever it is that we will call 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: \(\text{Corr}(\epsilon_xt, \epsilon_yt) = 0\)

Using the assumptions, we can rewrite (A2) as:

\begin{align}
\text{(A3)} \quad \text{Corr}[^x_t,^y_t] &= \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]+\text{Var}[\epsilon_xt]}(\text{Var}[y]+\text{Var}[\epsilon_yt])} \\
\text{or, multiplying by } \frac{\sqrt{\text{Var}[x]\text{Var}[y]}}{\sqrt{\text{Var}[x]\text{Var}[y]}} \\
\text{(A4)} \quad \text{Corr}[^x_t,^y_t] &= \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]\text{Var}[y]}(\text{Var}[x]+\text{Var}[\epsilon_xt])} \frac{\sqrt{\text{Var}[y]}}{\sqrt{\text{Var}[y]+\text{Var}[\epsilon_yt]}}
\end{align}

which is our equations (6). Rearranging terms and using the fact that \(\rho_{xy} = \text{Corr}(x, y) = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]\text{Var}[y]}}\) gives an expression for the real relatedness between $x$ and $y$:

\begin{align}
\rho_{xy} &= \rho_{sxt} \sqrt{\frac{\text{Var}[x]+\text{Var}[\epsilon_xt]}{\text{Var}[x]}} \frac{\sqrt{\text{Var}[y]+\text{Var}[\epsilon_yt]}}{\sqrt{\text{Var}[y]}} = \rho_{sxt} \sqrt{1 + \frac{\text{Var}[\epsilon_xt]}{\text{Var}[x]}} \sqrt{1 + \frac{\text{Var}[\epsilon_yt]}{\text{Var}[y]}}
\end{align}

This equation 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 importance of the measurement errors. To arrive at an estimate of this relative importance, we will assume that the error terms are uncorrelated through time.

Assumption 3a: \( \forall t, t': t \neq t' \rightarrow \text{Corr}(\varepsilon_{xt}, \varepsilon_{xt'}) = 0 \)

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

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

\[
\text{(A6)} \quad \text{Corr}[\hat{x}_t, \hat{x}_{t'}] = \frac{\text{Cov}[x + e_{xt} + e_{xt'}]}{\sqrt{\text{Var}[x + e_{xt}]\text{Var}[x + e_{xt'}]}}
\]

\[
\text{(A7)} \quad \text{Corr}[\hat{x}_t, \hat{x}_{t'}] = \frac{\text{Cov}[x] + \text{Cov}[x, e_{xt}] + \text{Cov}[e_{xt}, x] + \text{Cov}[e_{xt}, e_{xt'}]}{\sqrt{(\text{Var}[x] + 2\text{Cov}[x, e_{xt}])\text{Var}[x + e_{xt} + e_{xt'}]}}
\]

Using Assumptions 3a and 3b:

\[
\text{(A8)} \quad \text{Corr}[\hat{x}_t, \hat{x}_{t'}] = \frac{\text{Var}[x]}{\sqrt{\text{Var}[x] + \text{Var}[e_{xt}]}}
\]

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

Assumption 4a: \( \forall t, t': \text{Var}(\varepsilon_{xt}) = \text{Var}(\varepsilon_{xt'}) = \text{Var}(\varepsilon_x) \)

Assumption 4b: \( \forall t, t': \text{Var}(\varepsilon_{yt}) = \text{Var}(\varepsilon_{yt'}) = \text{Var}(\varepsilon_y) \)

\[
\text{(A9)} \quad \text{Corr}[\hat{x}_t, \hat{x}_{t'}] = \frac{\text{Var}[x]}{\text{Var}[x] + \text{Var}[e_x]}
\]

Similar derivations for \( y \) give:

\[
\text{(A10)} \quad \text{Corr}[\hat{y}_t, \hat{y}_{t'}] = \frac{\text{Var}[y]}{\text{Var}[y] + \text{Var}[e_y]}
\]

Using \( \rho \) for correlation, we arrive at (8) by substituting (Ag) and (A10) into (A5):

\[
\text{(A11)} \quad \rho_{xy} = \rho_{x_t y_t} \frac{1}{\sqrt{\rho_{x_t x_{t'}} \rho_{y_t y_{t'}}}} \frac{1}{\sqrt{\rho_{x_t x_{t'}} \rho_{y_t y_{t'}}}} = \frac{\rho_{x_t y_t}}{\sqrt{\rho_{x_t x_{t'}} \rho_{y_t y_{t'}}}}
\]
Appendix 2: Derivation of correction for attenuation bias based on T year average cross-type correlation

Let \( \bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t \) and \( \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \) 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:

\[
\text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[\frac{1}{T} \sum (x + \epsilon_{xt}), \frac{1}{T} \sum (y + \epsilon_{yt})]}{\sqrt{\text{Var}[\frac{1}{T} \sum (x + \epsilon_{xt})] \text{Var}[\frac{1}{T} \sum (y + \epsilon_{yt})]}} \tag{A12}
\]

\[
\text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[\sum (x + \epsilon_{xt}), \sum (y + \epsilon_{yt})]}{\sqrt{\text{Var}[\sum (x + \epsilon_{xt})] \text{Var}[\sum (y + \epsilon_{yt})]}} \tag{A13}
\]

Assuming constant measurement error variances over time (Assumptions 4a and 4b), the numerator of (A13) can be written as:

\[
\text{Cov}[\sum (x + \epsilon_{xt}), \sum (y + \epsilon_{yt})] = T^2 \text{Cov}[x, y] + T \sum \text{Cov}[x, \epsilon_{yt}] + T \sum \text{Cov}[\epsilon_{xt}, y] + \sum \sum \text{Cov}[\epsilon_{xt}, \epsilon_{yt}] \tag{A14}
\]

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

\[
\text{Cov}[\sum (x + \epsilon_{xt}), \sum (y + \epsilon_{yt})] = T^2 \text{Cov}[x, y]. \tag{A15}
\]

Expanding the first term of the denominator of (A13), we derive the following:

\[
\text{Var}[\sum (x + \epsilon_{xt})] = T^2 \text{Var}[x] + 2T \text{Var}[\epsilon_{xt}] + \sum \text{Cov}[\epsilon_{xt}, \epsilon_{xt}] + \sum \sum \text{Cov}[x, \epsilon_{xt}] \tag{A16}
\]

which under the before mentioned assumption simplifies to:

\[
\text{Var}[\sum (x + \epsilon_{xt})] = T^2 \text{Var}[x] + T \text{Var}[\epsilon_{x}] \tag{A17}
\]

Due to similar considerations for \( y \), the second term in the denominator of (A13) is:

\[
\text{Var}[\sum (y + \epsilon_{yt})] = T^2 \text{Var}[y] + T \text{Var}[\epsilon_{y}] \tag{A18}
\]

Substituting (A15), (A17) and (A18) into (A13), we get:

\[
\text{Corr}[\bar{x}, \bar{y}] = \frac{T^2 \text{Cov}[x, y]}{\sqrt{T^2 \text{Var}[x] + T^2 \text{Var}[\epsilon_{xt}] + 2T \text{Var}[\epsilon_{xt}] + \sum \text{Cov}[\epsilon_{xt}, \epsilon_{xt}] + \sum \sum \text{Cov}[x, \epsilon_{xt}]}} = \frac{\text{Cov}[x, y]}{\sqrt{\text{Var}[x] + \text{Var}[\epsilon_{xt}] + \text{Var}[y]}} \tag{A19}
\]

Rearranging (A9) yields the following expression for \( \text{Var}[\epsilon_{x}] \):

\[
\text{Var}[\epsilon_{x}] = \frac{\text{Var}[x]}{\text{Corr}[\bar{x}, \bar{y}]} - \text{Var}[x] = \frac{1 - \text{Corr}[\epsilon_{x}, \epsilon_{xt}]}{\text{Corr}[\epsilon_{x}, \epsilon_{xt}]} \text{Var}[x] \tag{A20}
\]

Substituting (A20) and its counterpart for \( y \) into (A19) gives:
\[
(A21) \quad \text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]+\frac{1}{\text{Var}[y]} \text{Corr}[x,y]^2}} \sqrt{\frac{1}{\text{Var}[x]+\frac{1}{\text{Var}[y]} \text{Corr}[x,y]^2}} \sqrt{\frac{1}{\text{Var}[y]}} \sqrt{\frac{1}{\text{Var}[y]}}
\]

Now we can rearrange the terms to arrive at:

\[
(A22) \quad \text{Corr}[\bar{x}, \bar{y}] = \frac{\text{Cov}[x,y]}{\sqrt{\text{Var}[x]\text{Var}[y]}} \left( \sqrt{1+\frac{1-\text{Corr}[x,y]^2}{\text{Corr}[x,y]^2}} \sqrt{1+\frac{1-\text{Corr}[y,y]^2}{\text{Corr}[y,y]^2}} \right)
\]

which can be rewritten in the form of (9):

\[
(A23) \quad \rho_{xy} = \rho_{\bar{x}\bar{y}} \sqrt{1+\frac{1-\rho_{xy}^2}{\rho_{\bar{x}\bar{y}}^2}} \sqrt{1+\frac{1-\rho_{yy}}{\rho_{\bar{y}\bar{y}}}}.
\]

Equation (A23) shows that, if we use the correlation between relatedness estimates that have been averaged across several years, the measurement error bias reflected in the year-on-year correlation disappears at a predictable rate as \( T \) grows sufficiently large. The reason is that, on average, measurement errors cancel out as long as they are truly uncorrelated. Therefore, if (A23) and (A11) give the same results, this validates the assumptions we use in the derivations above.
### Tables

#### Table 1: Occupations in labour market segments

<table>
<thead>
<tr>
<th>Occupations</th>
<th>Average # employees (2003-2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Management Occupations</strong></td>
<td>579,977</td>
</tr>
<tr>
<td>629: Foremen, master mechanics</td>
<td>114,882</td>
</tr>
<tr>
<td>722: Technical ships officers, ships engineers</td>
<td>4,405</td>
</tr>
<tr>
<td>751: Entrepreneurs, managing directors, divisional managers</td>
<td>333,014</td>
</tr>
<tr>
<td>761: Members of Parliament, Ministers, elected officials</td>
<td>2,215</td>
</tr>
<tr>
<td>762: Senior government officials</td>
<td>109,410</td>
</tr>
<tr>
<td>763: Association leaders, officials</td>
<td>16,053</td>
</tr>
<tr>
<td><strong>Sales Occupations</strong></td>
<td>1,671,452</td>
</tr>
<tr>
<td>681: Wholesale and retail trade buyers, buyers</td>
<td>397,942</td>
</tr>
<tr>
<td>682: Salespersons</td>
<td>888,757</td>
</tr>
<tr>
<td>683: Publishing house dealers, booksellers</td>
<td>21,281</td>
</tr>
<tr>
<td>687: Commercial agents, travelers</td>
<td>171,112</td>
</tr>
<tr>
<td>701: Forwarding business dealers</td>
<td>79,847</td>
</tr>
<tr>
<td>703: Publicity occupations</td>
<td>63,680</td>
</tr>
<tr>
<td>704: Brokers, property managers</td>
<td>8,579</td>
</tr>
<tr>
<td>705: Landlords, agents, auctioneers</td>
<td>27,635</td>
</tr>
<tr>
<td>836: Interior, exhibition designers, window dressers</td>
<td>12,619</td>
</tr>
<tr>
<td><strong>IT Occupations</strong></td>
<td>435,433</td>
</tr>
<tr>
<td>774: Data processing specialists</td>
<td>435,433</td>
</tr>
</tbody>
</table>

(…): 3-digit codes of the German Classifications of Occupations 1973
Table 2: Employment by labor market segment (2003-2006)

<table>
<thead>
<tr>
<th>Geography</th>
<th>High Wage Earners</th>
<th>Low Wage Earners</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td>10,601,365</td>
<td>10,469,329</td>
<td>10,173,929</td>
<td>10,219,006</td>
</tr>
<tr>
<td></td>
<td>High Wage Earners</td>
<td></td>
<td>9,533,278</td>
<td>9,375,802</td>
<td>9,062,606</td>
<td>9,059,832</td>
</tr>
<tr>
<td></td>
<td>Low Wage Earners</td>
<td></td>
<td>1,927,625</td>
<td>1,915,847</td>
<td>1,782,826</td>
<td>1,776,020</td>
</tr>
<tr>
<td></td>
<td>High Wage Earners</td>
<td></td>
<td>1,840,230</td>
<td>1,747,091</td>
<td>1,705,423</td>
<td>1,696,164</td>
</tr>
<tr>
<td></td>
<td>Low Wage Earners</td>
<td></td>
<td>8,486,815</td>
<td>8,386,286</td>
<td>8,162,923</td>
<td>8,152,740</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7,451,703</td>
<td>7,326,594</td>
<td>7,084,920</td>
<td>7,026,548</td>
</tr>
<tr>
<td>West-Germany</td>
<td>High Wage Earners</td>
<td></td>
<td>1,9805,105</td>
<td>19,470,691</td>
<td>18,829,493</td>
<td>18,744,137</td>
</tr>
<tr>
<td></td>
<td>Low Wage Earners</td>
<td></td>
<td>17,330,987</td>
<td>17,029,640</td>
<td>16,444,819</td>
<td>16,362,077</td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
<td>19,805,105</td>
<td>19,470,691</td>
<td>18,829,493</td>
<td>18,744,137</td>
</tr>
<tr>
<td></td>
<td>Local Movers</td>
<td></td>
<td>519,788</td>
<td>504,273</td>
<td>493,297</td>
<td>485,933</td>
</tr>
<tr>
<td></td>
<td>Non-Local Movers</td>
<td></td>
<td>1,445,658</td>
<td>1,431,396</td>
<td>1,388,718</td>
<td>1,386,246</td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td></td>
<td>363,839</td>
<td>364,861</td>
<td>366,079</td>
<td>374,009</td>
</tr>
<tr>
<td>Occupations</td>
<td>Sales</td>
<td></td>
<td>519,788</td>
<td>504,273</td>
<td>493,297</td>
<td>485,933</td>
</tr>
<tr>
<td></td>
<td>IT</td>
<td></td>
<td>1,445,658</td>
<td>1,431,396</td>
<td>1,388,718</td>
<td>1,386,246</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
<td>363,839</td>
<td>364,861</td>
<td>366,079</td>
<td>374,009</td>
</tr>
</tbody>
</table>
Table 3: Cross-industry labor flows for whole of Germany

<table>
<thead>
<tr>
<th>labor market segment</th>
<th>N</th>
<th>random flows</th>
<th>all</th>
<th>high wages</th>
<th>low wages</th>
</tr>
</thead>
<tbody>
<tr>
<td>different sector</td>
<td>81.3%</td>
<td>85.7%</td>
<td>60.4%</td>
<td>57.1%</td>
<td>63.4%</td>
</tr>
<tr>
<td>same sector</td>
<td>10.6%</td>
<td>7.8%</td>
<td>5.8%</td>
<td>5.8%</td>
<td>5.8%</td>
</tr>
<tr>
<td>same sub-sector</td>
<td>4.1%</td>
<td>2.7%</td>
<td>8.3%</td>
<td>9.1%</td>
<td>7.5%</td>
</tr>
<tr>
<td>same 2-digit industry</td>
<td>3.1%</td>
<td>2.7%</td>
<td>12.7%</td>
<td>13.5%</td>
<td>12.0%</td>
</tr>
<tr>
<td>same 3-digit industry</td>
<td>0.8%</td>
<td>0.7%</td>
<td>5.7%</td>
<td>6.4%</td>
<td>5.2%</td>
</tr>
<tr>
<td>same 4-digit industry</td>
<td>0.2%</td>
<td>0.3%</td>
<td>7.2%</td>
<td>8.1%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Total flows</td>
<td>3,385,391</td>
<td>1,618,221</td>
<td>1,772,674</td>
<td></td>
<td></td>
</tr>
<tr>
<td>labor market segment</td>
<td>GERMANY</td>
<td></td>
<td>GERMANY</td>
<td></td>
<td>OCCUPATIONAL GROUPS</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
<td>----------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>East</td>
<td>West</td>
<td>management</td>
<td>sales</td>
<td>IT</td>
</tr>
<tr>
<td>different sector</td>
<td>61.1%</td>
<td>61.4%</td>
<td>60.4%</td>
<td>53.9%</td>
<td>58.2%</td>
</tr>
<tr>
<td>same sector</td>
<td>8.0%</td>
<td>4.6%</td>
<td>8.4%</td>
<td>3.1%</td>
<td>2.0%</td>
</tr>
<tr>
<td>same sub-sector</td>
<td>10.2%</td>
<td>8.4%</td>
<td>10.9%</td>
<td>14.9%</td>
<td>27.1%</td>
</tr>
<tr>
<td>same 2-digit industry</td>
<td>13.7%</td>
<td>16.7%</td>
<td>14.6%</td>
<td>20.0%</td>
<td>10.7%</td>
</tr>
<tr>
<td>same 3-digit industry</td>
<td>7.0%</td>
<td>8.9%</td>
<td>5.7%</td>
<td>8.2%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Total flows</td>
<td>2,053,869</td>
<td>425,591</td>
<td>76,553</td>
<td>183,665</td>
<td>56,054</td>
</tr>
</tbody>
</table>

Local flows are defined as flows where the job switcher’s old and new work places are at most 100 km apart. All other flows are defined as non-local flows.
Table 5: Top 20 of industry pairs with the strongest skill-relatedness

<table>
<thead>
<tr>
<th>Rank</th>
<th>Industry of origin</th>
<th>Employment</th>
<th>Industry of destination</th>
<th>Employment</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6711: Administration of financial markets</td>
<td>653</td>
<td>6712: Security broking and fund management</td>
<td>633</td>
<td>9,925.3</td>
</tr>
<tr>
<td>2</td>
<td>1592: Production of ethyl alcohol from fermented materials</td>
<td>656</td>
<td>1591: Manufacture of distilled potable alcoholic beverages</td>
<td>1,727</td>
<td>7,483.6</td>
</tr>
<tr>
<td>3</td>
<td>1413: Quarrying of slate</td>
<td>562</td>
<td>1450: Other mining and quarrying n.e.c.</td>
<td>2,041</td>
<td>5,729.3</td>
</tr>
<tr>
<td>4</td>
<td>4512: Test drilling and boring</td>
<td>920</td>
<td>1120: Service activities incidental to oil and gas extraction, excluding surveying</td>
<td>1,067</td>
<td>4,399.5</td>
</tr>
<tr>
<td>5</td>
<td>1724: Silk-type weaving</td>
<td>552</td>
<td>1711: Preparation and spinning of cotton-type fibres</td>
<td>3,263</td>
<td>4,002.1</td>
</tr>
<tr>
<td>6</td>
<td>1725: Other textile weaving</td>
<td>2,395</td>
<td>1724: Silk-type weaving</td>
<td>552</td>
<td>2,998.0</td>
</tr>
<tr>
<td>7</td>
<td>2214: Publishing of sound recordings</td>
<td>1,587</td>
<td>2465: Manufacture of prepared unrecorded media</td>
<td>768</td>
<td>2,846.6</td>
</tr>
<tr>
<td>8</td>
<td>1120: Service activities incidental to oil and gas extraction, excluding surveying</td>
<td>1,067</td>
<td>4512: Test drilling and boring</td>
<td>920</td>
<td>2,545.4</td>
</tr>
<tr>
<td>9</td>
<td>1110: Extraction of crude petroleum and natural gas</td>
<td>1,361</td>
<td>1413: Quarrying of slate</td>
<td>562</td>
<td>2,485.5</td>
</tr>
<tr>
<td>10</td>
<td>125: Other farming of animals</td>
<td>736</td>
<td>123: Farming of swine</td>
<td>1,990</td>
<td>2,442.8</td>
</tr>
<tr>
<td>11</td>
<td>1110: Extraction of crude petroleum and natural gas</td>
<td>1,361</td>
<td>1120: Service activities incidental to oil and gas extraction, excluding surveying</td>
<td>1,067</td>
<td>2,437.4</td>
</tr>
<tr>
<td>12</td>
<td>2465: Manufacture of prepared unrecorded media</td>
<td>768</td>
<td>9212: Motion picture and video distribution</td>
<td>1,098</td>
<td>2,289.2</td>
</tr>
<tr>
<td>13</td>
<td>1715: Throwing and preparation of silk, including from noils, and throwing and texturing of synthetic or artificial filament yarns</td>
<td>1,743</td>
<td>1824: Manufacture of other wearing apparel and accessories n.e.c.</td>
<td>2,423</td>
<td>2,061.8</td>
</tr>
<tr>
<td>14</td>
<td>6523: Other financial intermediation n.e.c.</td>
<td>5,653</td>
<td>6712: Security broking and fund management</td>
<td>633</td>
<td>2,038.8</td>
</tr>
<tr>
<td>15</td>
<td>2232: Reproduction of video recording</td>
<td>503</td>
<td>9212: Motion picture and video distribution</td>
<td>1,098</td>
<td>1,995.2</td>
</tr>
<tr>
<td>16</td>
<td>2662: Manufacture of plaster products for construction purposes</td>
<td>642</td>
<td>1412: Quarrying of limestone, gypsum and chalk</td>
<td>1,683</td>
<td>1,971.8</td>
</tr>
<tr>
<td>17</td>
<td>1724: Silk-type weaving</td>
<td>552</td>
<td>1725: Other textile weaving</td>
<td>2,395</td>
<td>1,917.6</td>
</tr>
<tr>
<td>18</td>
<td>2665: Manufacture of fibre cement</td>
<td>1,156</td>
<td>2664: Manufacture of mortars</td>
<td>947</td>
<td>1,822.2</td>
</tr>
<tr>
<td>19</td>
<td>9212: Motion picture and video distribution</td>
<td>1,098</td>
<td>2232: Reproduction of video recording</td>
<td>503</td>
<td>1,769.7</td>
</tr>
<tr>
<td>20</td>
<td>6711: Administration of financial markets</td>
<td>653</td>
<td>6523: Other financial intermediation n.e.c.</td>
<td>5,653</td>
<td>1,630.7</td>
</tr>
</tbody>
</table>

Skill-relatedness is based on the entire German labor market.
### Table 6a: Correlation of transformed skill-relatedness estimates across time and labor market segments (4-digit level)

<table>
<thead>
<tr>
<th>Year</th>
<th>Germany high (1)</th>
<th>Germany low (2)</th>
<th>East-Germany high (1)</th>
<th>East-Germany low (2)</th>
<th>West-Germany high (3)</th>
<th>West-Germany low (4)</th>
<th>Geography local (1)</th>
<th>Geography non-local (2)</th>
<th>Occupations management (1)</th>
<th>Occupations sales (2)</th>
<th>Occupations IT (3)</th>
<th>Occupations others (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2004-05</td>
<td>0.51</td>
<td>0.44</td>
<td>0.40</td>
<td>0.36</td>
<td>0.49</td>
<td>0.43</td>
<td>0.48</td>
<td>0.46</td>
<td>0.38</td>
<td>0.41</td>
<td>0.32</td>
<td>0.49</td>
</tr>
<tr>
<td>2005-06</td>
<td>0.49</td>
<td>0.43</td>
<td>0.38</td>
<td>0.36</td>
<td>0.47</td>
<td>0.42</td>
<td>0.47</td>
<td>0.46</td>
<td>0.35</td>
<td>0.40</td>
<td>0.32</td>
<td>0.48</td>
</tr>
<tr>
<td>2006-07</td>
<td>0.48</td>
<td>0.43</td>
<td>0.38</td>
<td>0.35</td>
<td>0.47</td>
<td>0.42</td>
<td>0.46</td>
<td>0.45</td>
<td>0.37</td>
<td>0.39</td>
<td>0.30</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Depending on the exact combination of two relatedness types, the number of observations (i.e., industry combinations) ranges from 21170 to 248502.
### Table 6b: Correlation of transformed skill-relatedness estimates across time and labor market segments (2-digit level)

<table>
<thead>
<tr>
<th></th>
<th>Germany high (1)</th>
<th>East-Germany high (1)</th>
<th>West-Germany high (3)</th>
<th>Geography local (1)</th>
<th>Occupations management (1)</th>
<th>Occupations sales (2)</th>
<th>Occupations IT (3)</th>
<th>Occupations others (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2004-05</td>
<td>0.78</td>
<td>0.73</td>
<td>0.66</td>
<td>0.77</td>
<td>0.75</td>
<td>0.64</td>
<td>0.62</td>
<td>0.42</td>
</tr>
<tr>
<td>2005-06</td>
<td>0.76</td>
<td>0.72</td>
<td>0.66</td>
<td>0.74</td>
<td>0.74</td>
<td>0.62</td>
<td>0.57</td>
<td>0.40</td>
</tr>
<tr>
<td>2006-07</td>
<td>0.75</td>
<td>0.69</td>
<td>0.65</td>
<td>0.74</td>
<td>0.72</td>
<td>0.60</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>(1)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(2)</td>
<td>0.97</td>
<td>1.00</td>
<td>0.97</td>
<td>0.93</td>
<td>0.93</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(3)</td>
<td>0.91</td>
<td>0.86</td>
<td>1.00</td>
<td>0.91</td>
<td>0.86</td>
<td>0.84</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>(4)</td>
<td>0.89</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.84</td>
<td>0.86</td>
<td>0.86</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Depending on the exact combination of two relatedness types, the number of observations (i.e., industry combinations) ranges from 2070 to 3192.
### Table 7: Firm diversification events by skill-relatedness category

<table>
<thead>
<tr>
<th>relatedness category</th>
<th>% of industry combinations</th>
<th>NACE relatedness</th>
<th>Germany high</th>
<th>Germany low</th>
<th>East-Germany high</th>
<th>East-Germany low</th>
<th>West-Germany high</th>
<th>West-Germany low</th>
<th>Geography local</th>
<th>Geography non-local</th>
<th>Occupations management</th>
<th>Occupations sales</th>
<th>Occupations IT</th>
<th>Occupations others</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73.6%</td>
<td>451</td>
<td>187</td>
<td>189</td>
<td>336</td>
<td>326</td>
<td>167</td>
<td>195</td>
<td>164</td>
<td>184</td>
<td>327</td>
<td>80</td>
<td>84</td>
<td>169</td>
</tr>
<tr>
<td>2</td>
<td>20.5%</td>
<td>15</td>
<td>132</td>
<td>134</td>
<td>156</td>
<td>133</td>
<td>162</td>
<td>137</td>
<td>155</td>
<td>177</td>
<td>120</td>
<td>85</td>
<td>41</td>
<td>157</td>
</tr>
<tr>
<td>3</td>
<td>2.7%</td>
<td>39</td>
<td>40</td>
<td>39</td>
<td>46</td>
<td>29</td>
<td>34</td>
<td>44</td>
<td>39</td>
<td>59</td>
<td>53</td>
<td>37</td>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>2.7%</td>
<td>96</td>
<td>158</td>
<td>145</td>
<td>108</td>
<td>154</td>
<td>176</td>
<td>147</td>
<td>155</td>
<td>177</td>
<td>120</td>
<td>85</td>
<td>54</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>0.6%</td>
<td>46</td>
<td>121</td>
<td>131</td>
<td>132</td>
<td>113</td>
<td>105</td>
<td>119</td>
<td>127</td>
<td>90</td>
<td>54</td>
<td>52</td>
<td>19</td>
<td>120</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>647</td>
<td>638</td>
<td>638</td>
<td>622</td>
<td>622</td>
<td>638</td>
<td>638</td>
<td>647</td>
<td>647</td>
<td>554</td>
<td>334</td>
<td>218</td>
<td>638</td>
</tr>
<tr>
<td>in percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>73.6%</td>
<td>69.7%</td>
<td>29.3%</td>
<td>29.6%</td>
<td>54.0%</td>
<td>52.4%</td>
<td>26.2%</td>
<td>30.6%</td>
<td>25.3%</td>
<td>28.4%</td>
<td>59.0%</td>
<td>24.0%</td>
<td>38.5%</td>
<td>26.5%</td>
</tr>
<tr>
<td>2</td>
<td>20.5%</td>
<td>2.3%</td>
<td>20.7%</td>
<td>21.0%</td>
<td>7.4%</td>
<td>4.7%</td>
<td>5.3%</td>
<td>6.9%</td>
<td>6.0%</td>
<td>9.1%</td>
<td>9.6%</td>
<td>11.1%</td>
<td>9.2%</td>
<td>5.8%</td>
</tr>
<tr>
<td>3</td>
<td>2.7%</td>
<td>6.0%</td>
<td>6.3%</td>
<td>6.1%</td>
<td>7.4%</td>
<td>4.7%</td>
<td>5.3%</td>
<td>6.9%</td>
<td>6.0%</td>
<td>9.1%</td>
<td>9.6%</td>
<td>11.1%</td>
<td>9.2%</td>
<td>5.8%</td>
</tr>
<tr>
<td>4</td>
<td>2.7%</td>
<td>14.8%</td>
<td>24.8%</td>
<td>22.7%</td>
<td>17.4%</td>
<td>24.8%</td>
<td>27.6%</td>
<td>23.0%</td>
<td>24.0%</td>
<td>27.4%</td>
<td>21.7%</td>
<td>25.4%</td>
<td>24.8%</td>
<td>24.3%</td>
</tr>
<tr>
<td>5</td>
<td>0.6%</td>
<td>7.1%</td>
<td>19.0%</td>
<td>20.5%</td>
<td>21.2%</td>
<td>18.2%</td>
<td>16.5%</td>
<td>18.7%</td>
<td>19.6%</td>
<td>13.9%</td>
<td>9.7%</td>
<td>15.6%</td>
<td>8.7%</td>
<td>18.8%</td>
</tr>
</tbody>
</table>

First column shows share of all possible combinations of two industries in each relatedness category. For management relatedness and the East German relatedness measures, over 73.6% of all industry combinations had zero flows and were therefore assigned a skill-relatedness value of exactly zero. This prevented us from drawing a distinction between the first and second relatedness category in these cases. The values shown here represent totals for the combined first and second category.
Figures

"spurious" flows

"real" flows

Figure 1: Definition of spurious cross-establishment labor flows
Figure 2: Industry space of Germany at the 4-digit level of the wz03 classification (2003-2007)
Figure 3a: Industry space high wage segment (Germany, 2003-2007)

Figure 3b: Industry space low wage segment (Germany, 2003-2007)
Figure 4a: Industry space West-Germany (all wages, 2003-2007)

Figure 4b: Industry space East-Germany (all wages, 2003-2007)
Figure 5a: Industry space management occupations (Germany, 2003-2007)

Figure 5b: Industry space sales occupations (Germany, 2003-2007)
Figure 5c: Industry space IT occupations (Germany, 2003-2007)

Figure 5d: Industry space other occupations (Germany, 2003-2007)
Figure 6a: Industry space, local flows (Germany, 2003-2007)

Figure 6b: Industry space, non-local flows (Germany, 2003-2007)