

# PAPERS on Economics & Evolution



MAX-PLANCK-GESELLSCHAFT

# 0906

**Skill-relatedness and firm diversification**

**by**

**Frank Neffke  
Martin Svensson Henning**

The *Papers on Economics and Evolution* are edited by the Evolutionary Economics Group, MPI Jena. For editorial correspondence, please contact: [evopapers@econ.mpg.de](mailto:evopapers@econ.mpg.de)

ISSN 1430-4716

© by the author

Max Planck Institute of Economics  
Evolutionary Economics Group  
Kahlaische Str. 10  
07745 Jena, Germany  
Fax: ++49-3641-686868

## Skill-relatedness and firm diversification

Frank Neffke\* and Martin Henning\*\*

Version 1: May 2009

**Version 2: October 2010**

### ABSTRACT

According to the knowledge-based view of the firm, a firm's workforce is its most important resource, and firms often diversify into activities that allow them to leverage human resources. Human capital also represents a main asset for employees. When switching jobs, individuals are expected to remain in industries that value the skills that they have developed in their previous work. Based on this observation, this article develops theoretical arguments and a statistical method that uses cross-industry labor flows to assess the skill-relatedness between industries. Industries classified in different sectors of the economy sometimes exhibit strong skill-relatedness linkages. Also, *industry space*, i.e., the resulting network that connects industries with overlapping skill requirements, is highly predictive of diversification moves on the part of firms. Diversification is found to be over 100 times more likely to occur into industries that have ties to a firm's core activity in terms of skills than into industries that do not. This effect of skill-relatedness eclipses the effect of other types of relatedness, such as value chain linkages and classification-based relatedness.

Keywords: diversification; relatedness; human capital; labor flows; skills.

\* Department of Applied Economics, Erasmus University Rotterdam. PO Box 1738, NL-3000 DR Rotterdam, the Netherlands. [neffke@ese.eur.nl](mailto:neffke@ese.eur.nl)

\*\*Department of Economic History and Department of Human Geography, Lund University. PO Box 7083, 22007 Lund, Sweden. [martin@keg.lu.se](mailto:martin@keg.lu.se)

Financial support from Region Skåne, Näringslivsenheten, is gratefully acknowledged.

## INTRODUCTION

Skills acquired in one industry can often also be used in other industries. Consider, for example, Edmund Rumpler's 'Tropfenwagen' of 1923. With a drag coefficient of 0.28, it is sleek even by modern-day standards. Rumpler was originally trained as an automobile engineer but the drop-shaped Tropfenwagen was a product of his experience streamlining airplanes. Although car manufacturers worldwide were experimenting with aerodynamic designs at the time, advances were made particularly quickly in Germany. Ironically, what indirectly contributed to Germany's early lead in aerodynamic car engineering was the country's defeat in WWI. One of the provisions of the Treaty of Versailles demanded that Germany immediately cease building airplanes. As a result, the country's highly skilled aero-engineers suddenly lost their jobs. Many of them sought employment in the car industry, which they subsequently revolutionized with their skill at aerodynamic design (Spiegel, 2009). In the aftermath of WWI, former German airplane firms (*e.g.*, Hans Grade's Grade Automobilwerke AG) mimicked the behavior of these engineers and turned to automobile manufacturing as well. In strategic management, it is widely acknowledged that firms tend to diversify into activities that are related to their current core competences, but this insight has not yet been linked to the micro-level 'diversifications' that employees undertake when they switch jobs. We argue that what we observe in the early German car industry holds more generally and that inter-industry relatedness affects both employee decisions to participate in particular firms and firm decisions to diversify into particular industries. In fact, we show below that this observation can be used to construct a relatedness indicator that predicts the direction of corporate diversification in a substantially more accurate way than alternative measures by exploiting information from cross-industry labor flows.

Relatedness studies have a key position in corporate strategy research. The coherence of a firm in terms of the inter-relatedness of its activities is often considered an important factor in explaining long-term performance (Teece *et al.*, 1994). Moreover, relatedness has been shown to play a role in a wide range of business practices, prominent among which are corporate diversification (Wernerfelt, 1984; Chatterjee and Wernerfelt, 1991), make-or-buy decisions (Chang and Singh, 1999; Bryce and Winter, 2009), strategies for coping with information asymmetries in knowledge-intensive acquisitions (Coff, 1999) and alliance formation (Rothaermel and Boeker, 2008). In spite of the importance of the concept, however, the definition and modes of measuring relatedness are often surprisingly imprecise. One significant challenge is that if relatedness refers to the similarities between resources used in different industries, then relatedness must have as many facets as there are distinct types of resources. Quantifying all types of relatedness and assessing the relative importance of each is a Herculean task.

Our proposal is therefore to construct a relatedness measure that focuses on those resources that are most often credited for a firm's competitive advantage in the modern knowledge economy: the skills embedded in a firm's human capital.

The core assumption in this article is that if skilled individuals can move from one industry to another, then the production processes in both industries are likely to draw on similar skills and are related in this sense. Building on this assumption, for each pair of industries, we assess to what extent cross-industry labor flows are excessive compared to a well-defined baseline; we call the degree of excessiveness the skill-relatedness of the industries. Toward this end, labor flows among 415 different industries were derived from employment data on the roughly nine million individuals who were registered in Sweden between 2004 and 2007. Unlike many existing relatedness measures, our relatedness measure is not limited to the manufacturing sector but also covers service industries. The indicator often conforms to intuition but links industries differently than traditional industry classifications do. For example, we find strong links between manufacturing industries and concomitant wholesale and repair activities, which are classified as part of the service sector. Moreover, skill-relatedness reveals a complex web of inter-industry linkages that is unlike the rigidly nested hierarchy of a standard classification system. We refer to this network as *industry space*.<sup>1</sup>

The focus on shared resources differentiates our contribution conceptually from recently proposed relatedness indicators based on the co-occurrences of industries in corporate portfolios (Teece *et al.*, 1994; Neffke and Svensson Henning, 2008; Bryce and Winter, 2009). Whereas such co-occurrence studies assume that corporate portfolios are coherent and derive relatedness among industries *a posteriori*, our relatedness index can be determined *a priori*. We can therefore actually use it to test whether or not corporate portfolios expand coherently. In using skill-relatedness to analyze corporate diversification, we found that firms enter industries that are strongly skill-related to their core industries. In fact, we estimated that the probability that a firm will diversify into an industry that is strongly skill-related to its core activity is 123 times larger than the probability that it will diversify into unrelated industries. The effects of other relatedness types, such as value chain linkages, are negligible in comparison.

The following section provides theoretical arguments for the claim that skill-relatedness spans a conceptual space that consists of inter-connected industries that affects both individuals' decisions regarding their participation in industries and firm diversification behavior. This section draws on the resource-based view of the firm, the work of March and Simon and recent advances in labor economics.

---

<sup>1</sup> This term was chosen as an analog to Hidalgo *et al.*'s (2007) product space.

The third section is devoted to constructing the skill-relatedness index, and the fourth section presents empirical analyses of firm diversification. The final section discusses outcomes and limitations and provides recommendations for corporate management, policy-makers and future research.

## PRIOR LITERATURE AND THEORY

### *A resource-based perspective on related diversification*

The resourced-based view of the firm (RBV: Wernerfelt, 1984; Barney, 1991) has enriched our understanding of how and why certain resources, assets or capacities can become sources of sustained competitive advantage for firms (*e.g.*, Dierickx and Cool, 1989; Peteraf, 1993; Eisenhardt and Martin, 2000). The RBV framework has also guided research on which new markets firms diversify into and why. Penrose (1959) preceded this work, arguing that a firm has an incentive to grow as long as some of its resources are left idle in its current activities. Resources are distinct from the services a firm derives from them in two respects. First, with time, there is a steady increase in services that firms tend to extract from the same resources as awareness of the full potential of a resource grows through learning-by-doing. Whenever these additional services are difficult to sell or lease through market exchange, firms benefit from putting them to use themselves (Teece, 1982). Second, a single resource typically provides a variety of services. If not just the amount, but also the breadth of services a firm can derive from its resources increases, then this may justify the redeployment of resources to new activities (Penrose, 1959; Teece, 1982; Montgomery and Hariharan, 1991). However, adopting the conceptual apparatus of the RBV tradition, it emerges that, in order to be competitive on the new market, the excess resources must yield a competitive advantage in the new activities a firm undertakes (*e.g.*, Wernerfelt, 1984; Chatterjee and Wernerfelt, 1991).

Thus, a diversification strategy should focus on identifying new activities that require resources that a firm already possesses but that are currently underleveraged. We will refer to such diversification as *related diversification*. Early empirical evidence of the advantages of related diversification has been mixed (see Chatterjee and Wernerfelt, 1991). However, there is consensus regarding some issues since the advent of more advanced methods to establish the degree to which industries are related. For instance, scholars have shown that related diversification is more common than unrelated diversification (Farjoun, 1994; Chang, 1996; Fan and Lang, 2000; Lien and Klein, 2010), and diversification via internal development is often more related to the core activity of a firm than diversification by acquisition is (Chang and Singh, 1999; Bryce and Winter, 2009).

Regardless, from a theoretical perspective, mere resource similarity is insufficient for firms from different industries to successfully enter one another's markets. Similarities must exist between resources that yield sustained competitive advantages in both industries (Lien and Klein, 2008). Indeed, these resources must be valuable, rare, inimitable and non-substitutable (Barney, 1991). Therefore, relatedness measures should not measure mere economies of scope (Panzar and Willig, 1981) but economies of scope in critical resources (or what Dierickx and Cool (1989) refer to as critical or strategic asset stocks). An additional complication arises from the fact that relatedness may have as many distinct dimensions as there are strategically relevant resources. Industry relatedness has been estimated in a variety of ways, and different measures may reflect different types of resource relatedness. However, most current measures either do not necessarily capture the relatedness created by strategic resources or do so by definition without clarifying which types of resources are creating the relatedness.

There are three distinct approaches to measuring relatedness. First, scholars have relied on the structure of the standard industry classification system. In these systems, like the American Standard Industrial Classification (SIC) or the European Nomenclature générale des Activités économiques dans les Communautés Européennes (NACE) classification system, industries are catalogued into hierarchically nested categories. Each level is represented by a digit, with the first digit representing the broad economic sector to which an industry belongs, the second the sub-sector, the third the sub-sub-sector, and so on. Classification-based relatedness typically counts the number of (starting) digits that industries have in common. This method is straightforward to implement but there is little theoretical foundation for assuming that the hierarchical structure of industrial classification systems reflects the prevalence of scope economies among industries, let alone in terms of critical resources. However, a number of authors have successfully used classification-based relatedness to complement more sophisticated measures (Chang, 1996; Farjoun, 1998; Lee and Lieberman, 2010). Nevertheless, the meaning of classification-based relatedness is not entirely clear. Indeed, all three papers cited above provide different interpretations. Whereas Farjoun argues that classification-based relatedness captures relatedness that originates in physical resources, Lee and Lieberman use the classification system to distinguish between activities within and outside a firm's primary domain. Further, according to Chang, this system is used to differentiate between intensive and extensive search.

The second strategy for measuring relatedness takes as its point of departure the idea that different industries are found in the same portfolio due to the economies of scope between industries. Hence, relatedness can be inferred from the co-occurrence of industries in portfolios. Scholars have used industrial portfolios of firms or plants (Teece *et al.*, 1994; Neffke and Svensson Henning, 2008;

Bryce and Winter, 2009) or export profiles of countries (Hidalgo *et al.*, 2007) to identify the existence of economies of scope at the firm, plant or country level, respectively. An advantage of these indicators is that they cover many industries, especially in the manufacturing sector. Moreover, because they are based on micro-level portfolio choices, co-occurrence methods exploit collective knowledge about economies of scope that are apparently strategically relevant to individual actors. A major disadvantage, however, is that co-occurrence methods are essentially outcome-based, which means that these methods first *assume* that portfolios are coherent and then *infer* the implied relatedness of industries based on observed portfolios. Due to this *a priori* assumption that the activities contained within a portfolio are related, research on the coherence of corporate portfolios can become tautological. An additional disadvantage is that although co-occurrence-based relatedness indices may indeed identify industry combinations that feature economies of scope, they do not shed any light on the sources of these economies of scope. This ambiguity often makes it difficult to assess what type of relatedness has actually been quantified.

In the third approach to measuring relatedness, the indicators are input-based. Whereas co-occurrence methods focus on the consequences of economies of scope, resource-based relatedness indices concentrate on their origins. Resource-based indices aim to capture similarities in the resources used in different industries. For example, Fan and Lang (2000) concentrate on material resources that are transmitted via value-chain linkages, whereas in Jaffe (1989), Engelsman and Van Raan (1991) and Breschi, Lissoni and Malerba (2003), technological resources are investigated using patent analysis. Furthermore, Farjoun (1994) focuses on human capital and studies similarities between the occupational profiles of industries (see also Chang, 1996; Farjoun, 1998; Chang and Singh, 1999). Because resource-based measures identify underlying sources of economies of scope, it is possible to study corporate diversification without falling into the trap of circular reasoning. However, given that only relatedness between strategic resources matters, complications arise because not all resources are equally important in all industries. As a result, resource-based relatedness indicators are often biased toward specific types of industries. For instance, patent-based indicators only shed light on relatedness among technology and patent-intensive industries, and material-based indicators may be useful for studying the manufacturing sector but not the service sector.

The only resource-based indicator that does not suffer from any of these disadvantages is Farjoun's (1994) human capital-based indicator. This indicator is especially interesting because, regardless of the industry, knowledge today is identified as the prime resource of firms by proponents of the knowledge based view of the firm (Grant and Spender, 1996), and human resources are often

regarded as the ultimate repository of firm knowledge (Grant, 1996). In the traditional RBV, the importance of knowledge is acknowledged, if only because the sustained competitive advantage derived from critical physical resources that cannot be easily acquired in market transactions is often retraceable to the knowledge required to make productive use of them (Chang, 1996). Moreover, Porter (1987) identifies the opportunity for skill-transfer and skill-sharing as one of the most important prerequisites for the emergence of synergies between activities, and Farjoun (1994, p. 187) points out the pervasiveness of human capital in the economy, arguing that ‘human expertise can be viewed as an important *class* of resources (*e.g.*, like physical resources), since it is not merely a functional resource like R&D or marketing, but combines a wide range of value-added activities (*e.g.*, production, finance, marketing etc.).’

Farjoun’s focus on human capital inputs and human resource relatedness is therefore entirely justified, in our opinion, because human resources, in a broad sense, are the central resource expected to affect firm diversification behavior in today’s economies. Unfortunately, Farjoun was not able to directly observe the sharing of human capital across industries but instead proxied this factor by investigating to what extent industries employ workers in the same occupations. This strategy obscures the heterogeneity of individuals and their human capital within the same occupation. Moreover, not all employees will contribute equally to a firm’s competitive advantage. Therefore, we propose a different way to measure human capital relatedness, or ‘skill-relatedness’: the use of co-occurrence techniques to study inter-industry labor flows. In essence, we argue that the inter-connectedness of industries that guides corporate diversification strategies also affects cross-industry labor flows. Reversing the argument, this means that understanding the structure of cross-industry labor flows may help one to understand corporate diversification.

#### *Cross-industry labor mobility as an outcome of a decision to participate*

Cross-industry labor flows occur when individuals leave firms in one industry to start new jobs in another. Abstracting from the flows into and out of unemployment (and inactivity), labor flows can be understood as the aggregated outcome of employees’ *decisions to participate* in firms (March and Simon, 1958/1993, pp. 103-131). March and Simon argue that such decisions depend on two factors: the desirability of a job and the ease of movement to other jobs. The desirability of a job depends on how individuals weigh the disutility of their work at a firm against the benefits that they receive from that firm. The balance of contributions and benefits (or, in the words of March and Simon, the contribution-inducement balance) does not only take into account general working conditions (*e.g.*, the degree of



autonomy, corporate culture, and working hours) and monetary benefits (e.g., salaries and pension schemes), but also values them against the outside options that individuals may have. In other words, an individual will consider the opportunity costs of alternative employment options when evaluating the desirability of his or her current job.<sup>2</sup> The ease with which he or she can move to another job is dependent on the availability of such jobs. However, March and Simon stress that this availability is subjective: the mere existence of an alternative is insufficient because the alternative must also be visible to the individual.

This reasoning at the firm level can be generalized to the decision to leave an entire industry. The desirability and ease of movement to jobs in other industries depends on a number of industry-specific factors, prominent among which are the size and growth rates of the industries. However, the degree to which jobs outside the individual's current industry are visible to and available for the individual also depends on the match between the old industry and the new one. We argue that the most important aspect of this match is the degree to which industries have similar skill-requirements. However, we will begin by discussing the role of similarities between corporate cultures and of the overlap between the professional and social networks of different industries.

The decision to participate in a firm is influenced by the degree to which that firm's corporate culture matches the values of the individual in question. The initial sorting of individuals, but also the subsequent alignment of personal values with corporate culture (March and Simon, 1958/1993; Chatman, 1991) will create some homogeneity in values among the employees of a firm. Moreover, as shown by Chatman and Jehn (1994), corporate values are largely specific to each industry. As a result, some industries will be more 'value-compatible' than others, and people may prefer to move to industries with a corporate culture similar to that of their old firm.<sup>3</sup> For instance, people working in the private sector may be less inclined to switch to public sector jobs and *vice versa*.

Another factor that affects labor mobility is social networks. Social networks are an important channel through which people become informed about new job opportunities and thus improve the visibility of alternative employment. An extensive body of literature has brought together evidence of the importance of networks in helping people find jobs (Lin and Dumin, 1986; Wegener, 1991) and in helping firms find employees (Fernandez, Castilla and Moore, 2000). Given that a significant portion of a person's social network consists of the professional network that he or she developed during her or his

---

<sup>2</sup> In the following pages, we assume that all labor moves are voluntary, neglecting lay-offs. Under unemployment, opportunity costs include the value of being in a state of unemployment. Otherwise, the situation is quite similar.

<sup>3</sup> This is not to deny that people often leave precisely because of value conflicts with their employers. However, in the end, this should lead to better sorting and an increased homogeneity of values in a firm's workforce.

career, it is likely that social networks reflect the industries in which a person worked. Industries whose employees have overlapping social networks may therefore exhibit elevated labor flows among them.<sup>4</sup>

The third aspect of the similarities between industries that is relevant to understanding labor moves is the question of similarities between human capital requirements. Human capital research has a long tradition in the field of labor economics. However, labor economists have traditionally treated human capital as a quantity that can be measured using variables such as educational attainment and the number of years of experience. In contrast, skills are intuitively associated with qualitative characteristics; *which skills* an individual possesses is more important than *how many skills* he or she possesses. Notwithstanding the importance of formal education, many skills are acquired through imitation and learning-by-doing (Polanyi, 1959). As a result, individuals establish portfolios of skills as part of their work experience, and many of these skills are linked to individuals' jobs. Human capital is therefore often highly specific in a number of different respects. The early work on the subject by Becker (1964) highlights the firm specificity of human capital. However, more recent work by Neal (1995) and Parent (2000) shows that human capital is often specific not only to the firm but also to the industry.<sup>5</sup> Indeed, although human capital is not so completely generic as to be useful in all activities in the economy, the current consensus in the field of labor economics seems to be that human capital is also not completely specific to a particular job. This idea echoes Teece's (1982, p. 45) assertion that human capital is *fungible*; '[a] characteristic of organizational knowledge is that it is often fungible to an important degree. That is, the human capital inputs employed by the firm are not always entirely specialized to the particular products and services which the enterprise is currently producing.' The degree of specificity of human capital constrains people's labor mobility. The prospect of making a portion of their skills obsolete by changing jobs limits the desirability of such an action for workers. However, job switchers can reduce such human capital destruction by searching for employment in industries that value skill sets similar to those used in the industry in which they currently work. Moreover, it is likely that individuals are also more aware of job opportunities in such industries and that such opportunities are more visible to them.<sup>6</sup> As a consequence, cross-industry labor flows should

---

<sup>4</sup> Although much of the literature on social networks and labor markets has considered the relative importance of different types of ties (*e.g.*, stressing the role of weak ties (Granovetter, 1973)), it suffices here to note that the ways in which social networks in different industries overlap may explain relatively large cross-industry labor flows.

<sup>5</sup> In recent work, authors have stressed the occupational and task dimensions of human capital specificity. In fact, currently, a true task-based approach to human capital is emerging in labor economics (Autor, Levy and Murnane, 2003; Ingram and Neumann, 2006; Goos and Manning, 2007; Poleteav and Robinson, 2008; Gathmann and Schoenberg, 2010; Nedelkoska 2010).

<sup>6</sup> Equally importantly, employers prefer to hire new employees who possess skills that will be useful to them in their new jobs.

be stronger when human capital is more interchangeable between industries (*i.e.*, when they are skill-related).

### *Hypotheses*

As previously mentioned, of the three sources of cross-industry labor mobility patterns that we discussed, we expect the skill-match of industries to be dominant. Indeed, having the right skills should be a *sine qua non* for gaining employment in an industry.<sup>7</sup> We began this section by arguing that skill-relatedness plays an important role in diversification strategy. Therefore, our contention that this skill-relatedness also affects cross-industry labor flows leads to the following hypotheses:

*Hypothesis 1. Firms diversify into new activities that are closely skill-related to their present core activities.*

In fact, we have argued that human capital is the pivotal firm resource, and we therefore propose the following:

*Hypothesis 2. The effect of skill-relatedness on firm diversification is considerably larger than that of other types of relatedness.*

### **MEASURING SKILL-RELATEDNESS**

Based on the previous section, one can conclude that individuals changing jobs are likely to move between industries that are skill-related. By reversing this reasoning it is plausible that industries that exhibit large labor flows among them are likely to be skill-related. The main intuition behind the methodology we develop below is therefore that one can use information on labor flows to measure skill-relatedness. In this article, such information is taken from data on the 9 million people per year in the official registers of Sweden for 2004 to 2007. In this timeframe, each year, some 4.5 million people

---

<sup>7</sup> However, the different inter-industry matches are likely to overlap, and it is difficult to disentangle their respective effects. Social networks, for example, include former colleagues and classmates who may also have similar skills. It is therefore partly for ease of exposition that we refer to our labor-flow-based measure as an index of skill-relatedness, and the reader should bear in mind that it may also express similarities between corporate culture and overlapping social networks to some extent.

were active on the Swedish labor market, and approximately 280,000 of them changed jobs. In 77.2 percent of these job changes, people moved to another industry, creating cross-industry labor flows. Additional details are provided in Table A1 of the Appendix.

However, there are two reasons why we deem the raw sizes of such cross-industry labor flows ill-suited to measuring skill-relatedness. First, not all individuals are likely to exhibit strong *industry-specific* skills. In particular, for employees in low-skill areas, but also for people who largely make use of generic skills, movement between jobs may not be based primarily on skill-relatedness. In fact, due to the weaker presence of industry-specific skills, the effects of similarities in corporate culture or of overlapping social networks on labor flows may be comparatively strong for these groups. Therefore, we worry that such labor flows will add noise to our measures. Thus, reducing the presence of this type of labor in our data should reduce the noisiness of our skill-relatedness estimates. An industry is most likely to pay the highest salaries to people with the most critical skills in the industry. However, although we often find a large portion of managers to be among the most well paid employees, management skills may not be as industry-specific as other types of highly remunerated skills.<sup>8</sup> Based on this reasoning, we dropped the lowest paid half of the employees in each industry and all managers from our data. In the Appendix, we describe our measurement strategy in greater detail and show that the individuals that we selected to leave out of our analyses exhibit patterns of labor flows that indeed seem more randomly scattered across industries than those of the sample that we use for our skill-relatedness calculations. Furthermore, we excluded all individuals who move to newly established plants to avoid circular reasoning when we attempt to explain corporate diversification based on skill-relatedness.

In Table 1, we report raw labor flows for the individuals whom we selected for further study. The table shows the five industry combinations with the largest labor flows.

- Table 1 about here -

It seems plausible that the combinations in Table 1 are related industries. For instance, hotels use strategic resources similar to those of restaurants, and hospital and medical practice activities are obviously related as well. Moreover, three of the combinations of industries are in the same 3-digit class, and the remaining two are in the same 2-digit class. However, another salient feature of the

---

<sup>8</sup> We are not arguing that skill-relatedness does not play a role in job switches for these labor types. We simply believe that factors other than skill-relatedness play a *relatively* stronger role here, suggesting that we risk greater contamination of our measures if we include these labor types. We will return to this issue in discussing our findings.

industries in Table 1 is their size. Except for hotels, all of these industries rank among the top 10 industries in Sweden in terms of employment. Clearly, raw cross-industry labor flows also depend on industry size. It is therefore insufficient to determine whether or not a flow is large; we need to assess whether or not a flow is *exceptionally* large. In other words, we must construct a baseline that reflects our expectations about the size of a labor flow in an industry combination based on some general characteristics of the industries involved.

Two such characteristics are the size and growth of an industry, which determine the total possible supply of and demand for job switchers. Moreover, as indicated by March and Simon (1958/1993), jobs in large and growing industries are also more visible. An important factor that affects labor flows by determining the desirability of jobs is wages. All else being equal, industries that offer higher wages should attract larger labor flows, and low wages are an obvious reason for people to leave an industry. Ideally, we would therefore control for industry size and wage levels. As explained in the Appendix, standard co-occurrence tools only allow us to take size effects into account. However, in Neffke and Svensson Henning (2008), the baseline amount of co-occurrences between two industries is predicted using regression analysis. Using this approach, we constructed predicted labor flows that take into account a complete set of industry-specific variables. Next, by calculating the extent to which observed labor flows exceed predicted labor flows, we quantified the skill-relatedness of two industries as the *relative excess labor flow*.

Formally, let  $F_{ij}^{obs}$  denote the observed flow of individuals who move from industry  $i$  to industry  $j$ . Furthermore, let  $\hat{F}_{ij}$  be a prediction of this flow based on industry-level variables. Skill-relatedness is then simply defined as the ratio of these quantities:

$$(1) \quad SR_{ij} = \frac{F_{ij}^{obs}}{\hat{F}_{ij}}$$

If  $SR_{ij}$  equals 1, then industries  $i$  and  $j$  are unrelated. Values over 1 indicate skill-relatedness, whereas values under 1 indicate relative skill-dissimilarity. To determine  $\hat{F}_{ij}$ , we must select an appropriate regression model. Labor flows are non-negative and integer-valued, which suggests the use of count data regression models. Because labor flows are actually zero in the vast majority of industries, a standard Poisson model will not provide a good fit. Instead, we used a zero-inflated negative binomial regression (zinb) procedure. The Appendix provides further details, a discussion of the construction of the significance levels of the skill-relatedness index, and a descriptive analysis of the results.

*Industry space*

Using the procedure just described, we calculated a matrix of skill-relatedness indices for all  $415 \times 414 = 171,810$  combinations of 415 different industries at the 4-digit level in the European NACE classification. This matrix describes a network (depicted in Figure 1) in which industries are the nodes, and the links represent skill-relatedness. We will refer to this network as 'industry space'. The different colors and shapes of the nodes represent the broad sectors to which industries belong according to the industrial classification system. The graph depicts the 2.5 percent strongest skill-relatedness links that are significant at the 5 percent level. The position of the nodes was determined using a spring-embedded algorithm, which scatters industries across the entire plane in such a way that more closely related industries are located near one another.<sup>9</sup> Therefore, industries that are close to each other in industry space are usually strongly skill-related.

In general, nodes of the same color cluster together in industry space. For instance, the circles for industries in manufacturing (dark blue), utilities (purple) and construction (black) are mostly found on the right side of the network. Service industries of different types, represented by squares, are predominantly located on the left side of the network, with financial services (yellow) forming a close-knit cluster in the upper left corner. However, upon closer inspection, we find many exceptions to this pattern. For example, some business services are surrounded by manufacturing industries, which reflects strong skill-linkages among industries from both sectors. Similarly, some wholesale and retail activities are well connected to upstream manufacturing industries. The finding that industries are often skill-related to industries in different sectors supports our earlier criticism of classification-based relatedness measures. The discrepancies between skill-relatedness and traditional industry classifications become even clearer when we focus on a single industry. As an illustration, we have highlighted the network of industries connected to the pharmaceutical preparations industry (upper right corner). The pharmaceutical preparations industry is related to other industries in the chemical sector but also to R&D business services, specialized wholesale activities and the surgical equipment and orthopedic appliances industry, which is formally part of the instruments sector. Although these skill-linkages are intuitively very plausible, they would have remained unnoticed if we had relied only on classification codes.

- Figure 1 about here -

---

<sup>9</sup> This picture was generated using the *NetDraw* software package (Borgatti, 2002).

**EMPIRICAL ANALYSIS OF FIRM DIVERSIFICATION**

Thus far, we have described the skill-relatedness index and shown that skill-relatedness links industries that we intuitively expect to have similar skill-requirements. In this section, we test the predictive performance of the index using hypotheses regarding firm diversification as formulated in our theory section. Toward this end, we constructed a firm-level database by aggregating individuals into establishments and establishments into firms.<sup>10</sup> On average, there were over 425,000 firms in the Swedish economy during the period from 2004 to 2007. Approximately 8,000 firms consisted of more than one establishment. If we define firm expansion as an event in which an existing firm opens a new establishment, then there were 4,376 expansion events per year. Most expansion events simply increase production capacity for existing activities, but in 649 cases, a firm opened an establishment in a new industry and thereby further diversified its industrial portfolio. All of these diversification events involve setting up a new plant. We did not study acquisitions, which are typically in less related industries (Bryce and Winter, 2009), thereby limiting the analyses to diversification by internal development.

The hypotheses at the end of the second section focus on whether and in what sense the industries into which a firm diversifies are related to the firm's core activity. We define the core activity of a firm as the industry in which it employs the largest portion of its labor force. Each of the 649 diversification moves<sup>11</sup> can accordingly be regarded as a move *from* the core industry *into* a new industry.

Table 2 provides a summary of the observed diversification moves. The upper panel shows how close the new industries are to the firm's core activity in the industrial classification system (henceforth, their NACE relatedness to the core industry). Surprisingly, most diversifications cross the boundaries between 1-digit sectors. This finding shows that according to the industrial classification system, diversification predominantly takes place between unrelated industries. If we instead use the skill-relatedness index, then this finding must be revised substantially. In some 80 percent of all industry combinations, skill-relatedness equals zero. However, these industry combinations account for only 77 of the 649 diversification moves (or less than 12%).<sup>12</sup> The lower panel of Table 2 tabulates diversification moves for different skill-relatedness values. Unlike NACE relatedness, which can only assume four

---

<sup>10</sup> Statistics Sweden identifies firms and establishments using the so-called DEE system (Andersson and Arvidson, 2006). As a result, changes in the legal status of establishments or firms do not lead to changes in identifiers.

<sup>11</sup> This excludes five diversification moves for which industries were too small to calculate skill-relatedness.

<sup>12</sup> Moreover, in over two-thirds of these diversifications, skill-relatedness is not significantly different from 1, indicating that it is uncertain whether or not these diversifications are actually unrelated.

different values in a 4-digit classification system, skill-relatedness is a continuous measure. To make the two relatedness measures comparable, we transformed skill-relatedness into a categorical variable that assumes four different values such that each category contains the same number of industry combinations as the corresponding category of NACE-relatedness. Focusing on the lowest relatedness categories, we find that 66.4 percent of all diversification moves are unrelated according to the NACE system, whereas only 33.4 percent are according to skill-relatedness. In fact, except in the highest category, diversification moves between industry pairs in high skill-relatedness categories always substantially outnumber those in high NACE-relatedness categories.

- Table 2 about here -

### *Diversification probabilities*

A more rigorous way to study the predictive power of skill-relatedness is to focus on the choices that a diversifying firm faces. Imagine a firm  $f$  that is active in  $N_f$  industries and has decided that it wants to diversify. In principle, there are  $415 - N_f$  possible diversification candidates. These candidate industries can be represented by the set  $X_f = \{x_1, x_2, \dots, x_{415-N_f}\}$  that contains all industries in the economy but excludes the original industrial portfolio of firm  $f$ . Let us furthermore assume that industry  $i \notin X_f$  is the core industry of firm  $f$ . We can now describe the situation in which firm  $f$  diversifies into industry  $j$  using the following three arrays:

$$(7) \quad C_f = \begin{bmatrix} (i, x_1) \\ (i, x_2) \\ \vdots \\ (i, j) \\ \vdots \\ (i, x_{415-N_f}) \end{bmatrix}; \quad R_f = \begin{bmatrix} RSR(i, x_1) \\ RSR(i, x_2) \\ \vdots \\ RSR(i, j) \\ \vdots \\ RSR(i, x_{415-N_f}) \end{bmatrix}; \quad d_f = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$

$C_f$  contains all possible diversification moves by firm  $f$  from its core industry  $i$  into new industries.  $R_f$  is a vector of skill-relatedness indices corresponding to the industry combinations listed in  $C_f$ . Finally, the vector  $d_f$  tags diversification moves that have taken place with a one and contains zeros elsewhere.<sup>13</sup> By

<sup>13</sup> Note that although in this example,  $d_f$  contains zeros in all but one position (which corresponds to  $C_f(i, j)$ , where it assumes a value of 1, indicating that  $f$  diversifies into industry  $j$ ), a firm that enters more than one industry will have several elements in  $d_f$  equal to 1.



stacking the above vectors for all diversifying firms, we obtained a database that summarizes the 649 diversification moves together with the alternative diversification options that existed but were not chosen in the vectors  $C$ ,  $R$  and  $d$ .<sup>14</sup>

We used this database to estimate the diversification probabilities for different ranges of skill-relatedness values<sup>15</sup> and plotted the outcome in Figure 2 together with a confidence band of plus or minus two standard deviations.<sup>16</sup>

- Figure 2 about here -

Although the absolute diversification probability is not very informative,<sup>17</sup> the increase in diversification probability associated with moving to higher skill-relatedness values is striking. When we move from industry combinations for which skill-relatedness equals zero to industry combinations for which skill-relatedness values are around 50, we observe a 123-fold increase in the probability of diversification. We regard this leap in diversification probabilities as strong support for the hypothesis that firms diversify into new activities that are closely skill-related to their present core activities (Hypothesis 1).

Hypothesis 2 states that skill-relatedness should be a more important factor in diversification moves than other types of relatedness. Table 2 has already shown that skill-relatedness outperforms NACE-relatedness. However, some diversification moves may be motivated by a desire for vertical integration. A third form of relatedness that we would therefore expect to play an important role in diversification decisions is value chain linkages. Toward this end, we constructed two variables based on the Swedish input-output table for 2005.<sup>18</sup>

$$\text{InputRelatedness}(i, j) = \frac{IN(i, j)}{\sum_i IN(i, j)};$$

$$\text{OutputRelatedness}(i, j) = \frac{OUT(i, j)}{\sum_j OUT(i, j)}.$$

<sup>14</sup> Formally, the sample suffers from selection bias. Because we only used information on those firms that diversify, we neglect the fact that there may be a number of other firms that undertook diversification efforts but failed. However, we have no way of determining whether or not firms undertook such efforts except for the fact that some of them succeeded.

<sup>15</sup> We calculate the relative frequency of 1s in  $d_f$  for 15 different intervals of the values in  $R_f$ .

<sup>16</sup> To calculate the confidence bands of a given interval, we assume that the values in  $d_f$  are drawn from a binomial distribution with a probability of success equal to the relative frequency of diversification and the number of draws equal to the number of elements in the  $R_f$  interval.

<sup>17</sup> Given that there are 409.8 candidate industries for a firm on average and that firms diversify into 1.16 industries at a time on average, the average diversification probability should be  $1.16/409.8 = 0.28$  percent.

<sup>18</sup> This information is taken from the National Accounts, Statistics Sweden, [www.scb.se](http://www.scb.se).

where

$IN(i, j)$ : the value of inputs that were sourced from industry  $i$  by industry  $j$ ; and

$OUT(i, j)$ : the value of output that was sold by industry  $i$  to industry  $j$ .

These variables express the relative importance of one industry as a buyer or supplier of another industry's goods and services. Because input-output tables are only available at the 2-digit level, we must assume that the input and output relatedness values at the 2-digit level are representative of the 4-digit level as well.

#### *Regression analysis*

To assess the relative importance of each relatedness type, we moved to a multivariate setting, using our diversification dummy  $d$  as a dependent variable in logit regressions. Table 3 summarizes the outcomes of these analyses. Apart from point estimates and standard errors, we also report effect sizes to facilitate the interpretation of the coefficients of the logit equation. The effect sizes represent the change in diversification probability associated with a given increase in the variable at hand while keeping all other variables at their median values. For the relatedness variables, which were transformed into categorical variables, we report the rise in predicted diversification probability associated with a move from the lowest to the highest category. For the employment variables, effect sizes refer to an increase of half a standard deviation above their median values.

- Table 3 about here -

In the first column, diversification is regressed on skill-relatedness only. The estimated effect of skill-relatedness is highly significant. Moreover, the effect sizes show that when moving from low to high categories, the expected diversification probability increases by about 10.8 percent. This increase is of the same order of magnitude as in Figure 2, which provides further support for Hypothesis 1.

Column (2) shows the parameter estimates when we control for overall Swedish employment in the origin and destination industries.<sup>19</sup> As expected, larger industries are associated with higher diversification probabilities. However, the effect of skill-relatedness remains unchanged when we add these control variables. In columns (3) and (4), variables for NACE and value chain relatedness are added to the regression equation. Although they are often highly significant, the effects of these variables are

---

<sup>19</sup> To reduce skewness, we log-transformed these variables.

minimal (around 0.1%).<sup>20</sup> Their inclusion, however, does reduce the size of the effect of skill-relatedness (by 36% in column (3) and 44% in column (4)). Nonetheless, the effect of skill-relatedness remains strong. Given that the estimated diversification probability for industry combinations in the lowest skill-relatedness category is only 0.061 percent, an effect of 6.2 percent represents a 102-fold increase in diversification probability. Furthermore, the skill-relatedness effect eclipses the effects of all of the other relatedness variables. We therefore conclude that there is strong support for Hypothesis 2: skill-relatedness and the fungibility of human capital play an extremely important role in corporate diversification decisions, also when compared to the effect of other types of relatedness.

#### *Additional analyses: skill-combinations*

Thus far, we have used skill-relatedness to quantify to what extent new industries draw on sets of skills similar to those of a firm's core industry. However, firms that are already diversified do not simply use one skill set. Instead, they combine the skill sets of a number of different industries. Since Schumpeter's (1951) suggestion that innovations are 'neue Kombinationen' (*i.e.*, hitherto untried combinations of existing ideas), the notion of combining skills and knowledge has been a central theme in theories of discovery and novelty creation. Given that diversification is a form of innovation from the firm perspective (*i.e.*, the activity is new to the firm, though it may not be new to the world), the combination of skill sets that a firm possesses may also matter for diversification. In fact, some authors propose that to enter a new industry, a firm must possess the right *combination* of skills (*e.g.*, Farjoun, 1998).

We investigate the role of combinations of skill sets in the model in column (5) of Table 3. Referring to industries as skill-related if their skill-relatedness index does not fall into the lowest relatedness category (which contains 85.7% of all industry combinations), we counted the number of industries in a firm's portfolio to which a candidate industry is related. This number indicates to what extent a candidate industry is related not only to a firm's core activity but to its full range of activities. Column (5) adds two variables to the regression equation: the number of related industries in the portfolio and the total number of industries in the portfolio. The reported effect sizes of these variables correspond to a move from their lowest to their highest observed values in the sample. At first sight, the introduction of these variables lowers the effect of skill-relatedness substantially. However, the new variables should be regarded as decomposing the skill-relatedness variable into two components: a pure

---

<sup>20</sup> Skill-relatedness is also superior in statistical terms. Estimating an ANOVA for the model in column (4), we find that the share of explained variance that can be attributed to skill-relatedness is 2.3 times higher than the combined share of the remaining three relatedness variables.

skill-component and a skill-combination component. Thus, the decrease in effect size does not invalidate our earlier findings but it results from disentangling different aspects of skill-relatedness. Both, the relatedness to a firm's core activity and the relatedness to its broader portfolio of activities, account for a substantial increase in diversification probabilities. It appears that the likelihood that a firm will diversify into a particular industry depends not only on whether the firm's core industry is skill-related to this industry but also on what number of the firm's current industries are related to the industry.

### DISCUSSION AND CONCLUSIONS

Our empirical analyses demonstrate the strong predictive validity of the labor flow-based relatedness index for firm diversification. The effects of skill-relatedness are both statistically significant and substantial in an economic sense. Diversification probabilities increase over 100-fold when we move from low to high levels of skill-relatedness. Further, this effect is over 63 times as strong as the effect of output relatedness, which has the second highest impact. Hence, one important and novel finding of our study is that value chain linkages are far less important than human capital similarities in firm diversification moves. Furthermore, we provide powerful evidence that industrial classification systems are ill-suited to measuring relatedness. Skill-relatedness often crosses the boundaries between broadly defined sectors such as manufacturing and services, and after we correct for skill-relatedness, the effect of NACE relatedness on firm diversification is negligible. One caveat, however, is that we only considered diversification via internal development. Diversification through acquisitions lies beyond the scope of this article but certainly merits closer attention. Similarly, the refocusing of firm activities through divestitures is an important topic for future work that will help us to gain a fuller understanding of the evolution of firm portfolios.

#### *Cross-industry labor flows*

The use of labor flows to quantify skill-relatedness also requires further qualification, as we have already argued. Labor flows are not only driven by similarities in terms of skill requirements but also by compatibility between corporate cultures and social networks. Thus, our relatedness index may capture more than simply skill-relatedness. For instance, although the five industries that are most skill-related to social work activities are all health care industries, small-scale farming and the activities of religious organizations are also highly ranked. It is plausible that labor flows to these industries are caused more by compatible values than by compatible skills. Similarly, the public administration industries form a

rather tight but peripheral cluster in industry space, and this may reflect differences between the corporate culture of the private and the public sector rather than skill-relatedness. Regardless, our overall impression after a close inspection of the skill-relatedness matrix is that skill similarities dominate industry space linkages. However, because compatibility between corporate culture and social networks may also affect the diversification strategies of firms, our empirical analyses suffer from some degree of observational equivalence with respect to these alternative interpretations.

A similar problem arises because people are often reluctant to migrate. Cross-industry labor flows will therefore also be affected by the geographical distribution of industries. There is no straightforward solution to this problem. Because sharing a common pool of labor is an important reason for firms to cluster geographically (see Rosenthal and Strange (2004) for an in-depth discussion of such agglomeration externalities), industries that use similar skills will often be found in the same region. Controlling for geographical variables in the prediction of labor flows is therefore likely to do more harm than good. One could, however, restrict the skill-relatedness index to those labor flows that involve people who switch industries and, at the same time, move from one living space to another. However, given the limited size of the Swedish population, such a strategy was unfeasible in this study.

Experimenting with different labor flows may also prove fruitful in other respects. For instance, we dropped management occupations from our data based on the idea that management skills are not necessarily industry-specific. However, it would also be possible to verify this by isolating the cross-industry labor flows of managers and using them to construct a measure of management skill-relatedness. In fact, the methodology we developed in this article can be used to estimate a broad range of different relatedness indices, including marketing relatedness (based on occupations involved in marketing and sales) or technological relatedness (based on labor flows of engineers and technicians), as long as there is a sufficiently large sample of individuals. Moreover, in approximately one-third of all diversification moves, we found no evidence of skill-relatedness between industries. One reason for this may be that skill-relatedness is simply not always measured with sufficient precision. However, not all diversification efforts will be aimed at leveraging human resources. A closer inspection of such skill-unrelated diversification could yield important insights regarding reasons for diversification that are not resource-based.

#### *Novelty creation*

In the additional analyses that we presented, we have shown that the number of activities that a firm combines affects the probability of diversification into a particular candidate industry. More specifically,

we showed that it is the number of activities a firm engages in that are related to the new industry that matters (and not the sheer number of activities). One explanation for this finding can be found in the literature on innovation, which suggests that combining skills from different contexts is often frustrated by a lack of common ground between contexts (Nootboom 2000). On this basis, Nootboom proposes that most novelty creation arises from interactions between people whose knowledge bases are sufficiently different to offer the necessary scope for exchanging ideas but sufficiently similar to ensure mutual understanding. In other words, novelty arises by combining ideas from contexts at an optimal cognitive distance from one another. Because skill-related industries are different in some capacity but are still sufficiently similar for people to be able to shift between them, they can be thought of as contexts with a favorable cognitive distance, which may explain the dominant role of skill-relatedness in the diversification decisions of firms.

Along this line of reasoning, one may speculate that firms benefit from building a portfolio of skill-related activities, although we did not investigate this directly. For instance, rotating personnel within such a portfolio may offer both the firm and its employees a constant flow of novel perspectives without compromising the integrity of the organization as a whole. Indeed, Nonaka, Toyama and Nagata (2000, p. 3) argue that ‘... individuals and organizations have a potential to *grow together* through the process of knowledge creation’ (italics in original). Knowledge is created by putting information into a particular context, and firms provide individuals with ‘ba,’ or a ‘shared context in motion’ (Nonaka, Toyama and Nagata, 2000, p. 8), in which they cooperate to generate new interpretations and knowledge. Diversification widens this context and thus induces novelty creation. Because some industries can more easily be put into an overarching context than others, a strategy of skill-related diversification may offer the required change without disrupting the shared context.<sup>21</sup> Thus, the changing position of a firm’s portfolio of activities in industry space may affect its innovative capabilities. We believe that the study of firms and their portfolios from this perspective constitutes a promising new avenue for future research.<sup>22</sup>

---

<sup>21</sup> Nootboom’s (2000, pp. 183-188) cycle of discovery is relevant in this context as well, as one could argue that firms can use subsequent moves into related industries to maintain a cycle of discovery by using their existing knowledge in new but related domains.

<sup>22</sup> A particularly interesting aspect of skill-relatedness in this respect is that it is essentially asymmetrical in nature. In other words, the skill-relatedness from  $i$  to  $j$  is not necessarily the same as the skill-relatedness from  $j$  to  $i$ . Such asymmetries may indicate that the skills involved in certain industries have different levels of complexity, which may lead to irreversibilities in portfolio development.

*The ontological status of industry space*

An important point to note is that we have not linked the diversification strategy of firms directly to the skills of their employees. On the contrary, because we excluded individuals who work in new plants from the labor flows, labor flows and diversification events were effectively derived from independent datasets. Instead of directly linking the micro-level of individual skills to the macro-level of corporate competences, we thus introduced an intermediate level of analysis: the industry space network.<sup>23</sup> We then proposed that this intermediate level affects both the micro-level decisions by individuals seeking to participate in firms and the macro-level selection of diversification candidates by firms. Our empirical analyses strongly support the idea that this intermediate level does connect corporate diversification patterns to patterns of cross-industry job switching. To theoretically justify endowing the industry space dimension with its own ontological position, we can extend a line of argument presented by Chang (1996). Chang (p. 588) claims that ‘... the list of lines of business, is the ‘state’ description of the firm’s knowledge base embedded in the individual lines’ routines.’ Thus, one way to think of industry space is as reflecting the structure through which the knowledge or skills associated with individual industries are linked.

However, it remains unclear what ontological status we should attribute to industry space. One might initially think of industry space as being pre-determined by nature: which industries’ skill-bases can be connected might depend, for example, on the laws of physics. However, which skill-bases have actually been linked in human knowledge depends on humanity’s understanding of nature, implying that industry space changes as technological frontiers are pushed back. Yet, this is unlikely to be the full story. Social constructivist theories of knowledge draw attention to the fact that the actual organization of knowledge reflects social processes. Industry space could thus also be regarded as a social artifact, allowing, for instance, cultural idiosyncrasies to play a role and industry space to differ across countries. The introductory anecdote about aerodynamic car design in Germany supports such a perspective, showing that ceasing airplane manufacturing caused a forced dialogue between the car and airplane industries and led to new mental maps that inspired novel car designs. This line of reasoning also opens up the possibility that firms construct their own internal industry spaces, connecting activities that other firms may find difficult to connect<sup>24</sup> and echoing Nonaka *et al.*’s (2000) idea that firms create ‘contexts-in-motion.’ It may also be relevant to dynamic capabilities (Teece, Pisano and Shuen, 1997) because the capacity of firms to address novel situations may rely (in part) on their ability to bend the currently

---

<sup>23</sup> In this respect, skill-relatedness is no different from other relatedness indices.

<sup>24</sup> This is one way to interpret Apple’s recent success at activities that are atypical for a computer manufacturer.

dominant industry space within the organization. Given that micro-level labor market databases are becoming available for an increasing number of countries, future research on the dynamics of and cross-country differences within industry space could clarify its ontological status and may thereby uncover hitherto unexplored aspects of the process of discovery.

*Implications for management and policy making*

The main lesson of our study for management is that managers must be aware of the latent potential for diversification in their current workforce. Skill-related industries offer opportunities to diversify without compromising corporate coherence. Moreover, such moves may benefit both a firm and its employees because the concomitant favorable cognitive distances between activities increase the scope for learning. As skill-relatedness changes over time, new opportunities may constantly arise due to a sudden rewiring of industry space. Recognizing such new opportunities early may give firms an edge over their competitors. The value of skill-relatedness and industry space research is, however, not limited to firm-level management. It may also prove a powerful tool for understanding a local economy as a whole. Regions and countries are constantly exposed to the forces of structural change. This process of creative destruction destroys opportunities in one part of the economy while creating them in other parts. Some skills will be rendered obsolete, whereas other skills will become scarce, with tremendous losses in human capital investments as a result. Understanding the latent potential for diversification in the workforce may thus be as important for regional and national policy-makers as it is for corporate management.



## REFERENCES

- Andersson J, Arvidson G. 2006. Företagens och arbetsställes dynamik (FAD). Memo, Statistics Sweden.
- Autor DH, Levy F, Murnane RJ. 2003. The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics* **118**(4): 1279-1333.
- Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of Management* **17**(1): 99-120.
- Becker G. 1964. *Human Capital, a theoretical and empirical analysis with special reference to education*. Columbia University Press: New York.
- Borgatti SP. 2002. *NetDraw. Graph Visualization Software*. Analytic Technologies, Harvard.
- Breschi S, Lissoni F, Malerba F. 2003. Knowledge-relatedness in firm technological diversification. *Research Policy* **32**: 69-87.
- Bryce DJ, Winter SG. 2009. A general inter-industry relatedness index. *Management Science* **55**: 1570-1585.
- Chang SJ. 1996. An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance 1981-89. *Strategic Management Journal* **17**:587-611.
- Chang SJ, Singh H. 1999. The impact of modes of entry and resource fit on modes of exit by multibusiness firms. *Strategic Management Journal* **20**: 1019-1035.
- Chatterjee S, Wernerfelt B. 1991. The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal* **12**(1): 33-48.
- Chatman JA. 1991. Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly* **36**(3): 459-484.
- Chatman JA, Jehn KA. 1994. Assessing the relationship between industry characteristics and organizational culture: How different can you be? *Academy of Management Journal* **37**(3): 522-553.
- Coff RW. 1999. How buyers cope with uncertainty when acquiring firms in knowledge-intensive industries: Caveat emptor. *Organization Science* **10**(2): 144-161.
- Dierickx I, Cool K. 1989. Asset stock accumulation and sustainability of competitive advantage. *Management Science* **35**(12): 1504-1511.
- Eisenhardt KM, Martin JA. 2000. Dynamic capabilities: What are they? *Strategic Management Journal* **21**: 1105-1121.
- Engelsman EC, Van Raan AFJ. 1991. *Mapping of Technology, A First Exploration of Knowledge Diffusion Amongst Fields of Technology*. Policy Studies on Technology and Economy (BTE) Series, The Hague.
- Fan JPH, Lang LHP. 2000. The Measurement of Relatedness: An application to Corporate Diversification. *Journal of Business* **73**(4): 629-660.
- Farjoun M. 1994. Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups. *Organization science* **5**(2): 185-199.
- Farjoun M. 1998. The independent and joint effects of the skill and physical bases of relatedness in diversification. *Strategic Management Journal* **19**(7): 611-630.
- Fernandez RM, Castilla EJ, Moore P. 2000. Social capital at work: Networks and employment at a phone center. *American Journal of Sociology* **105**: 1288-1356.
- Gathmann C, Schönberg U. 2010. How general is human capital? A task-based approach. *Journal of Labor Economics* **28**(1): 1-49.
- Goos M, Manning A. 2007. Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics* **89**(1): 118-133.
- Granovetter MS. 1973. The strength of weak ties. *American Journal of Sociology* **78**(6): 1360-1380.
- Grant RM. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* **17**: 109-122.

- Grant RM, Spender J-C. 1996. Knowledge and the firm: overview. *Strategic Management Journal* **17**: 5-9.
- Hidalgo CA, Klinger B, Barabási A-L, Hausmann R. 2007. The Product Space Conditions the Development of Nations. *Science* **317**: 482-487.
- Ingram BF, Neumann GR. 2006. The returns to skill. *Labor Economics* **13**: 35-59.
- Jaffe A. 1989. Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy* **18**: 87-97.
- Lee GK, Lieberman MB. 2010. Acquisition vs. internal development as modes of market entry. *Strategic Management Journal* **31**: 140-158.
- Lien BL, Klein PG. 2009. Using competition to measure relatedness. *Journal of Management* **35**(4): 1078-1107.
- Lin N, Dumin M. 1986. Access to occupations through social ties. *Social networks* **8**(4): 365-385.
- Long JS. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage: Thousand Oaks.
- March JG, Simon HA. 1958. *Organizations*. Second edition (1993). Blackwell Publishers: Cambridge.
- Montgomery CA, Hariharan S. 1991. Diversified expansion by large established firms. *Journal of Economic Behavior and Organization* **15**: 71-89.
- Neal D. 1995. Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics* **13**(4): 653-677.
- Nedelkoska L. 2010. *Human capital in transition: On the changing skill requirements and skill transferability*. Doctoral thesis, Friedrich-Schiller-Universität Jena, Germany.
- Neffke F, Svensson Henning M. 2008. Revealed Relatedness: Mapping Industry Space. Working paper, Papers in Evolutionary Economic Geography, Urban and Regional research centre Utrecht.
- Nonaka I, Toyama R, Nagata A. 2000. A firm as a knowledge creating entity: A new perspective on the theory of the firm. *Industrial and Corporate Change* **1**: 1-20.
- Nooteboom B. 2000. *Learning and Innovations in Organizations and Economics*. Oxford University Press: Oxford.
- Panzar JC, Willig RD. 1981. Economies of scope. *American Economic Review* **71**(2): 268-272.
- Parent D. 2000. Industry-specific capital and the wage profile: Evidence from the National Longitudinal Survey of Youth and the Panel of Income Dynamics. *Journal of Labor Economics* **18**(2): 306-323.
- Penrose ET. 1959. *The Theory of the Growth of the Firm*. Blackwell: Oxford.
- Peteraf MA. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal* **14**: 179-191.
- Polanyi M. 1959. *The Tacit Dimension*. Chicago University Press: Chicago.
- Poletaev M, Robinson C. 2008. Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys. *Journal of Labor Economics* **26**(3): 387-420.
- Porter ME. 1987. From competitive advantage to corporate strategy. *Harvard Business Review* **65**(3): 43-59.
- Rosenthal SS, Strange WC. 2004. Evidence on the Nature and Sources of Agglomeration Economies. In: *Handbook of Urban and Regional Economics* (4). Henderson J, Thisse JF (eds.). Elsevier: Amsterdam.
- Rothaermel FT, Boeker W. 2008. Old technology meets new technology: complementarities, similarities and alliance formation. *Strategic Management Journal* **29**: 47-77.
- Schumpeter JA. 1951. *The theory of economic development. An inquiry into profits, capital, credit, interest and the business cycle*. Harvard University Press: Cambridge (MA). Original edition first published 1911.
- Spiegel online*. 2009. Stromlinien-Show, Geheimnis des rasenden Eis. 21 January.
- Teece DJ. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization* **3**: 39-63.

- Teece DJ, Pisano G, Shuen A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* **18**(7): 509-533.
- Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding corporate coherence. Theory and evidence. *Journal of Economic Behavior and Organization* **23**: 1-30.
- Wegener B. 1991. Job mobility and social ties: Social resources, prior job, and status attainment. *American Sociological Review* **56**(1): 60-71.
- Wernerfelt B. 1984. A resource-based view of the firm. *Strategic Management Journal* **5**: 171-180.

**APPENDIX: CROSS-INDUSTRY LABOR FLOWS AND ESTIMATION OF SKILL-RELATEDNESS**

In this Appendix, we describe how we constructed the matrix of skill-relatedness indices. This matrix, together with the STATA and MATLAB code used in the article will be available on the first author's website.

*Data and labor flows*

The original datasets were provided by Statistics Sweden and contain employment and wage information on the roughly 9 million individuals registered in Sweden during the 2004-2007 period. There were over 4.5 million active individuals on the Swedish labor market at that time. By our definition, a person is active on the labor market if he or she: (a) has a non-zero wage income and (b) works in a plant with a registered industry code. Plants are assigned to 4-digit industries according to the SNI2002 industry classification system. At this level of aggregation, the classification is the same as the European NACE (Rev 1.1) classification. All 4-digit industries that employ fewer than 250 persons per year on average were omitted because their labor flows are too small. The resulting database contains a total of 415 different industries. Furthermore, to avoid reverse causality issues, we omitted all individuals who move to plants that did not exist in 2004, the first year for which we have data.

Labor flows consist of the sum total of individual labor market moves. We registered a change in employment as a labor market move if an employee changed jobs from one year to the next by moving to another plant at another firm. By requiring that an employee change firms and establishments, we avoided the possibility that a large number of individuals switched industries due to the reclassification of plants. On average, nearly 600,000 individuals (about 12.5% of the active labor force) changed jobs each year. However, slightly over half of them moved to newly established plants and were thus omitted from the analysis. Table A1 summarizes information on all individuals working in plants that existed prior to 2005.

- Table A1 about here -

Of the 280,000 remaining job switchers, almost 65,000 changed industries. As discussed in the theory section, we can attain the most accurate possible picture of skill-relatedness by limiting our analysis to individuals who are likely to possess specialized industry-specific skills. One strategy is to use information on the occupations of individuals. However, such an approach would be plagued by problems. The process of selecting specific 'skill-intensive' occupations would be subjective at best and

arbitrary at worst. For instance, consider bookkeepers. In most firms, bookkeepers are viewed as support staff. However, in some business services, bookkeeping is a core competence. Leaving out all bookkeepers would be advisable in the former case but detrimental in the latter. Instead of identifying all individuals who *do* use important industry-specific skills, we therefore settled for a more modest aim and attempted to identify those individuals who are likely *not* to use many such skills in their work. We then excluded them from our analysis.

Ultimately, we excluded two types of individuals: (apparently) low-skilled individuals and individuals in management positions. In general, firms will reward employees most if they provide skills that confer competitive advantage to the firm. Individuals with few skills that are deemed critical in the industry will earn wages that are low relative to that industry's general wage level. We therefore disregarded all flows involving individuals who earn wages lower than their industry's median wages. However, employees in management positions are also likely to be among the high wage earners. Management skills are often rather generic. Therefore, we also excluded all individuals in management positions.<sup>25</sup> We want to stress, however, that we do not believe that all individuals who earn low wages have low skill levels or that managers do not have any industry-specific skills. Our contention is simply that these groups may have skills that erode less when switching industries and that their labor flows are less indicative of skill-relatedness on average. By leaving them out of our analyses, we should therefore reduce the noise in our relatedness indices.

In Table A1, we find support for this claim. The individuals in our selection show evidence of markedly different behavior from low-wage earners and managers in a way that suggests that their human capital is indeed less job-specific. For instance, low-wage earners are more mobile than the average individual; they are about twice as likely to change jobs as the individuals we selected. Moreover, both managers and low-wage earners more frequently move into industries in a radically different part of the economy, according to the industrial classification system. At the extreme, managers move to jobs in completely different sectors of the economy 20 percent more often than individuals in our sample. For low-wage earners the figure is even 50 percent higher.

#### *Estimating skill-relatedness*

As noted in the main text, labor flows do not only depend on skill-relatedness but also on certain general characteristics of the industries involved. This presents us with a problem similar to that in co-occurrence analysis: certain industries are overrepresented in any portfolio, regardless of their

---

<sup>25</sup> We first exclude all individuals in management occupations and then establish the median wage in an industry.

relatedness to other industries. Co-occurrence methods solve this problem by comparing raw co-occurrence counts to a baseline. Although the exact type of baseline differs from one author to another, the vast majority of articles use a baseline that reflects the different sizes of industries.<sup>26</sup>

However, there are also other industry characteristics that increase the visibility and desirability of jobs, and only controlling for size effects does not do justice to them. For this reason, we use a method outlined in Neffke and Svensson Henning (2008) that compares observed co-occurrences against a baseline generated using regression analysis. First, observed co-occurrences (or in our case, observed cross-industry flows) are used as a dependent variable in a regression model that uses all  $415 \times 414$  combinations of industries as observations. The vast majority of industry combinations exhibit no labor flows at all, which suggests that a zero-inflated negative binomial (zinb) model is appropriate. The zinb regression equation has two components: a regime selection equation and a count data component. The regime selection equation determines whether there will be any flow at all. Next, the count data component estimates the size of the flows, assuming that a non-zero regime is selected. The general zinb regression model can be written as follows:<sup>27</sup>

$$(2) \quad E(F_{ij}|v_i, w_j, \varepsilon_{ij}) = [1 - \pi_0(\gamma + v_i' \delta_i + w_j' \delta_j)] f(\alpha + v_i' \beta_i + w_j' \beta_j + \varepsilon_{ij}, \sigma)$$

In (2),  $\pi_0(\cdot)$  is the probability that a flow can, in principle, take place as a function of the vectors  $v_i$  and  $w_j$ . These vectors contain industry-level variables.  $f(\alpha + v_i' \beta_i + w_j' \beta_j + \varepsilon_{ij})$  describes the size of a labor flow as a function of the same industry-level variables. Using this model, we then estimate the constants  $\gamma$  and  $\alpha$  and the coefficient vectors  $\delta_i$ ,  $\delta_j$ ,  $\beta_i$ , and  $\beta_j$ . The variables that we use to construct our baseline predictions  $\hat{F}_{ij}$  are total employment, employment growth and average wages.<sup>28</sup>

<sup>26</sup> For instance, Teece *et al.* (1994) and Bryce and Winter (2009) compare their co-occurrence counts to what should be expected if the number of firms that own plants in an industry is kept constant, but plants are randomly reshuffled across firms. Hidalgo *et al.* (2007) divide observed co-occurrences by the total number of times the larger of the two product classes is found in any co-occurrence.

<sup>27</sup> The precise mathematical representation of the zinb model (*e.g.*, Long, 1997) is rather cumbersome. The probability that one will be able to observe any flow is modeled by a logistic CDF. The size of the flow is next modeled by what is, in essence, very similar to an exponential function.

<sup>28</sup> It is technically possible to add more variables. For instance, average educational attainment or the age composition of an industry may affect labor flows. However, these variables do not represent incentives for labor market moves so much as characterize the individuals who work in an industry. Consequently, they may proxy other characteristics of individuals, such as their skills, and including them as regressors may bias skill-relatedness estimates.

Our primary goal in the regression analyses is not causal analysis. Instead, we aim to arrive at the best possible prediction of labor flows based on the information in the regressors. We pool all data by summing labor flows and employment data across all available years to improve the efficiency of the estimates. After some experimentation, a model that uses variables in levels for the regime selection equation and log-transformed variables for the count data equation proves to perform best. We then estimate the following model:

$$(3) \quad E(F_{ij}|v_i, w_j, \varepsilon_{ij}) = [1 - \pi_0(\gamma + \delta_i emp_i + \delta_j emp_j)] \cdot \\ f(\alpha + \beta_{1i} \log(emp_i) + \beta_{2i} \log(wage_i) + \beta_{3i} growth_i \\ + \beta_{1j} \log(emp_j) + \beta_{2j} \log(wage_j) + \beta_{3j} growth_j)$$

with

- $emp_i$ : sum of employment in industry of origin  $i$  across 2004, 2005 and 2006;  
 $emp_j$ : sum of employment in destination industry  $j$  across 2005, 2006 and 2007;  
 $wage_i$ : average wage in industry of origin  $i$  across 2004, 2005 and 2006; and  
 $wage_j$ : average wage in destination industry  $j$  across 2005, 2006 and 2007.

Employment growth is defined as follows to create a more symmetrically distributed variable:

$$growth_i = \frac{emp_i(2006) - emp_i(2004)}{emp_i(2006) + emp_i(2004)} \\ growth_j = \frac{emp_i(2007) - emp_i(2005)}{emp_i(2007) + emp_i(2005)}$$

with the year to which the data refer in parentheses.

The outcomes of the regression are summarized in Table A2. The sizes of the origin and destination industry have the expected positive effects on the size of labor flows.<sup>29</sup> Fast-growing industries have both larger outflows and inflows of labor, which may reflect the greater volatility of these industries. Wage levels have no effect on the outflow of labor, but in line with the conjecture that high-wage industries offer desirable jobs, they do lead to a significant increase in labor inflow.

- Table A2 about here -

<sup>29</sup> The regime selection equation is framed in such a way that the probability of observing a flow of 0 is coded as a 1. Therefore, a negative effect in this equation indicates a positive effect of the variable on the dependent variable.

Using the point estimates of the parameters in equation (3), we calculate the expected labor flows in all industry combinations based on industry size, growth and wage information. For instance, the expected labor flow from hotels to restaurants is 7.7. However, we observe a labor flow of 892. Therefore, the estimated skill-relatedness from hotels to restaurants equals  $\frac{892}{7.7} = 115.3$ . This number is substantially larger than 1, which indicates that hotels and restaurants are strongly skill-related industries. In fact, a value of 115.3 puts these industries in the top 0.06 percent of all industry combinations in terms of skill-relatedness. If we repeat these calculations for all industry combinations, we obtain a matrix that consists of 171,810 different skill-relatedness indices.

#### *Determining the significance levels of skill-relatedness estimates*

As noted above, there are no labor flows between the vast majority of industries. As a result, the skill-relatedness matrix contains 81.3 percent elements equal to zero. However, in many such cases we would also predict labor flows to be negligible because, for instance, one of the industries involved is very small. In industry combinations with values for  $\hat{F}_{ij}$  that are only a fraction of 1, an increase in the labor flow from zero to one individual will lead to large changes in the skill-relatedness index. The corresponding relatedness indices should thus be interpreted with caution. The basic problem is that skill-relatedness is not estimated with equal precision for all industry combinations. Therefore, we construct confidence intervals.

Labor flows can be regarded as the outcome of job-switching decisions. To be more specific, we assume that all employees in an industry  $i$  have the option of switching to a new job in a new industry. For the sake of simplicity, we also assume that all of these individuals make their choices independently from one another and abstract from the fact that in reality, we can observe only one labor move per year for each individual. On this basis, each individual faces 415 independent choices. One is to stay in his or her current industry; the other 414 choices represent moves into each of the remaining 414 industries.

A job switch choice can then be modeled as a Bernoulli experiment with a probability of success equal to  $p_{ij}$ . At the aggregate level, the labor flow from  $i$  to  $j$ , or  $F_{ij}$ , is the outcome of a binomial experiment  $BIN(n, p)$  where  $n$  is equal to employment in industry  $i$ , and  $p$  is equal to  $p_{ij}$ :

$$(4) \quad F_{ij} \sim BIN(emp_i, p_{ij})$$



The question of how informative a specific labor flow is can now be rephrased as the question of how likely it is that one will observe  $F_{ij}^{obs}$  merely by chance. Let  $\hat{p}_{ij}$  be the expected counterpart of  $p_{ij}$ :

$$(5) \quad \hat{p}_{ij} = \frac{F_{ij}}{emp_i}$$

If we take  $\hat{p}_{ij}$  as a benchmark, then the question above translates into a statistical test of whether  $F_{ij}^{obs}$  is exceptional, assuming that  $\hat{p}_{ij}$  represents the real probability that an individual will move from industry  $i$  to industry  $j$ . If skill-relatedness exceeds one, then the one-sided test calculates the probability that  $F_{ij} \geq F_{ij}^{obs}$ , assuming that  $F_{ij}$  is generated via a  $BIN(emp_i, \hat{p}_{ij})$  process. The p-value of this test is calculated as follows:

$$(6) \quad P(x \geq F_{ij}^{obs} | p_{ij} = \hat{p}_{ij}) = 1 - \sum_{r=0}^{F_{ij}^{obs}-1} [\hat{p}_{ij}^r \cdot (1 - \hat{p}_{ij})^{emp_i-r} \binom{emp_i}{r}]$$

For skill-relatedness values smaller than 1, we reverse the inequalities:

$$(7) \quad P(x \leq F_{ij}^{obs} | p_{ij} = \hat{p}_{ij}) = \sum_{r=0}^{F_{ij}^{obs}} [\hat{p}_{ij}^r \cdot (1 - \hat{p}_{ij})^{emp_i-r} \binom{emp_i}{r}]$$

### Outcomes

Based on a p-value of 5 percent, skill-relatedness is significant and larger than 1 in 6,894 industry combinations (4.0% of all possible industry combinations). We find significant skill dissimilarity between industries in 2.8 percent of all cases. In 93.2 percent of all industry combinations, industries are neither significantly skill-related nor significantly skill-dissimilar. Most of the industries in these combinations are very small. In fact, when we weigh the industries of origin and destination by employment, significant estimates account for 56.4 percent of all industry combinations in the Swedish economy. Table A3 summarizes these findings.

- Table A3 about here -

## TABLES AND FIGURES

Table 1. Top five labor flows between 2004 and 2007.

Labor flow	Industry of origin	# employees	Industry of destination	# employees
1,301	8511: Hospital activities	101,810	8512: Medical practice activities	27,245
1,085	8531: Social work activities with accommodation	99,642	8511: Hospital activities	102,463
1,021	5530: Restaurants	25,959	5510: Hotels	13,232
933	8512: Medical practice activities	28,044	8511: Hospital activities	102,463
892	5510: Hotels	13,193	5530: Restaurants	24,585

Labor flow is the cumulative flow of labor from the industry of origin to the destination industry between 2004 and 2007. # of employees refers to the average annual number of employees during the periods 2004-2006 (industry of origin) and 2005-2007 (destination industry).

Table 2. Diversification moves by NACE- and skill-relatedness categories.

	# diversifications	% diversifications
Within the same 3-digit industry	46	7.1%
Within the same 2-digit industry	96	14.8%
Within the same 1-digit industry	76	11.7%
Between different 1-digit industries	431	66.4%
Within SR category 3 combinations	38	5.9%
Within SR category 2 combinations	127	19.6%
Within SR category 1 combinations	267	41.1%
Within SR category 0 combinations	217	33.4%

Skill-relatedness (SR) categories are defined in such a way that each category contains as many industry combinations as the corresponding row in the upper panel.

Table 3. Logit regression of firm diversification moves.

Variable	(1)		(2)		(3)		(4)		(5)	
	Point estimate	Effect size	Point estimate	Effect size	Point estimate	Effect size	Point estimate	Effect size	Point estimate	Effect size
SR	1.501*** (0.041)	0.108	1.779*** (0.052)	0.110	1.596*** (0.060)	0.070	1.562*** (0.060)	0.062	1.497*** (0.062)	0.046
NACE relatedness			0.331*** (0.052)	0.001	0.203*** (0.057)	0.001	0.203*** (0.057)	0.001	0.281*** (0.057)	0.001
Input relatedness					0.129 (0.069)	0.000	0.129 (0.069)	0.000	0.096 (0.070)	0.000
Output relatedness					0.318*** (0.066)	0.001	0.318*** (0.066)	0.001	0.271*** (0.064)	0.001
# ind in PF							0.027***		0.027***	0.001
# related ind in PF							0.673***		0.673***	0.015
log(employment(i))			0.288*** (0.028)	0.000	0.270*** (0.028)	0.000	0.233*** (0.028)	0.000	0.094*** (0.036)	0.000
log(employment(j))			0.642*** (0.027)	0.001	0.614*** (0.028)	0.001	0.575*** (0.029)	0.001	0.563*** (0.029)	0.000
Constant	-6.60*** (0.055)		-15.75*** (0.421)		-15.27*** (0.416)		-14.60*** (0.425)		-13.16*** (0.450)	
log-likelihood	-3985.3		-3632.0		-3612.9		-3593.7		-3525.4	
Nobs	230,755		230,755		230,755		230,755		230,755	

Significance levels: \*\*\*,  $p < 0.01$ ; \*\*,  $p < 0.025$ ; \*,  $p < 0.05$ . All relatedness variables are categorical. The limits on categories were chosen in such a way that the number of industry combinations within each category is as close as possible to the categories for the NACE relatedness measure. Effect size indicates the increase in diversification probability associated with moving from the minimum to the maximum value in the sample for the relatedness measures and for the count of (related) industries, and with moving up one standard deviation for the log(employment(*i*)) variables while all other variables are maintained at their median values.

Table A1. Labor flows in Sweden in 2004-2007.

	Full sample	Low wage earners	Managers	Selected sample
<b>General</b>				
<b>Labor force in existing plants</b>	<b>3,637,649</b>	<b>1,745,215</b>	<b>198,335</b>	<b>1,719,459</b>
% of population	(40.2%)	(19.3%)	(2.2%)	(19.0%)
<b>Labor flow</b>	<b>283,156</b>	<b>125,290</b>	<b>7,645</b>	<b>76,916</b>
% of active labor force	(7.8%)	(7.2%)	(3.9%)	(4.5%)
<b>Labor flow within</b>				
<b>same 4-digit industry</b>	<b>64,595</b>	<b>22,131</b>	<b>1,946</b>	<b>25,358</b>
% of total labor flow	(22.8%)	(17.7%)	(25.5%)	(33.0%)
<b>same 3-digit industry</b>	<b>78,514</b>	<b>27,327</b>	<b>2,376</b>	<b>30,128</b>
% of total labor flow	(27.7%)	(21.8%)	(31.1%)	(39.2%)
<b>same 2-digit industry</b>	<b>105,022</b>	<b>38,296</b>	<b>3,101</b>	<b>37,777</b>
% of total labor flow	(37.1%)	(30.6%)	(40.6%)	(49.1%)
<b>same 1-digit industry</b>	<b>138,438</b>	<b>52,621</b>	<b>4,079</b>	<b>46,601</b>
% of total labor flow	(48.9%)	(42.0%)	(53.4%)	(60.6%)
<b>Labor flow between</b>				
<b>different 1-digit industries</b>	<b>144,718</b>	<b>72,669</b>	<b>3,566</b>	<b>30,316</b>
% of total labor flow	(51.1%)	(58.0%)	(46.6%)	(39.4%)

Figures are yearly averages for the 2004-2007 period. Individuals in plants established after 2004 were omitted. Columns: <full sample>: all individuals in Sweden; <low-wage earners>: all individuals earning below the industry's median wage in consecutive years; <managers>: all individuals in management occupations in consecutive years; and <selected sample>: all individuals who in consecutive years are not in management occupations and earn at least the industry's median wage.

Table A2. Zero-inflated negative binomial regression of labor flows.

	Parameter estimate	Standard error	p-value
<i>Count data equation</i>			
log(emp_o)	0.851***	(0.006)	0.000
log(emp_d)	0.676***	(0.007)	0.000
growth_o	-1.318***	(0.141)	0.000
growth_d	0.363**	(0.161)	0.024
log(wage_o)	0.303***	(0.032)	0.000
log(wage_d)	0.679***	(0.032)	0.000
constant	-26.575***	(0.471)	0.000
<i>Regime selection equation</i>			
emp_o	1.24E-06***	(3.87E-07)	0.001
emp_d	-4.22E-04***	(2.12E-05)	0.000
constant	0.240***	(0.057)	0.000
<i>Over-dispersion parameter</i>			
log(alpha)	1.138***	(0.011)	0.000
Nobs	171,810		
Nobs flow=0	139,624		

\*, p<0.05; \*\*, p<0.025; and \*\*\*, p<0.01. emp\_o: sum of employment figures in industry of origin during the 2004-2006 period; emp\_d: sum of employment figures in destination industry during the 2005-2007 period; wage\_o: average yearly wage in industry of origin during the 2004-2006 period; wage\_d: average yearly wage in destination industry during the 2005-2007 period; growth\_o growth in industry of origin between 2004 and 2006 as defined in the Appendix; and growth\_d growth in destination industry between 2005 and 2007 as defined in Appendix.

Table A3. Frequency of significant skill-related industry combinations.

		Significant	Insignificant
<i>Level</i>			
$F_{ij}^{obs} = 0$	<b>dissimilar</b>	2,681	136,943
$F_{ij}^{obs} > 0$	<b>related</b>	6,894	15,168
	<b>dissimilar</b>	2,180	7,944
<i>Percentage</i>			
$F_{ij}^{obs} = 0$	<b>dissimilar</b>	1.56%	79.71%
$F_{ij}^{obs} > 0$	<b>related</b>	4.01%	8.83%
	<b>dissimilar</b>	1.27%	4.62%
<i>Employment-weighted</i>			
$F_{ij}^{obs} = 0$	<b>dissimilar</b>	10.65%	17.93%
$F_{ij}^{obs} > 0$	<b>related</b>	14.32%	9.30%
	<b>dissimilar</b>	31.47%	16.34%

Significance level: 5%. Weights:  $emp_i \cdot emp_j$ .

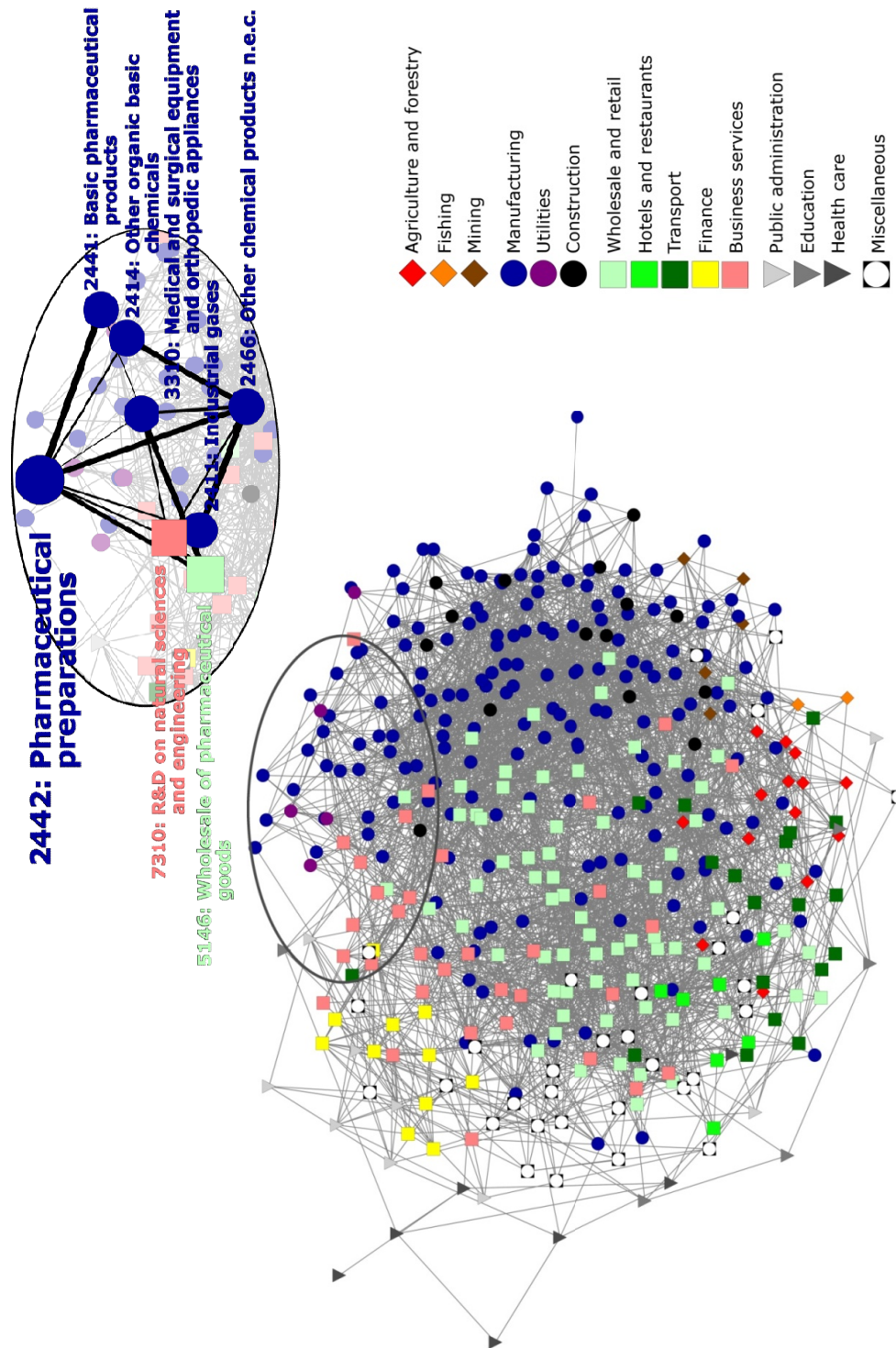


Figure 1. Industry space.

The figure depicts the strongest 2.5% of significant skill-relatedness links. The ego network for pharmaceutical preparations is highlighted in the upper right corner. This shows that pharmaceutical preparations are mainly linked to other chemical industries (24) but also to instruments industries (33), R&D business services (73) and wholesale activities (51).

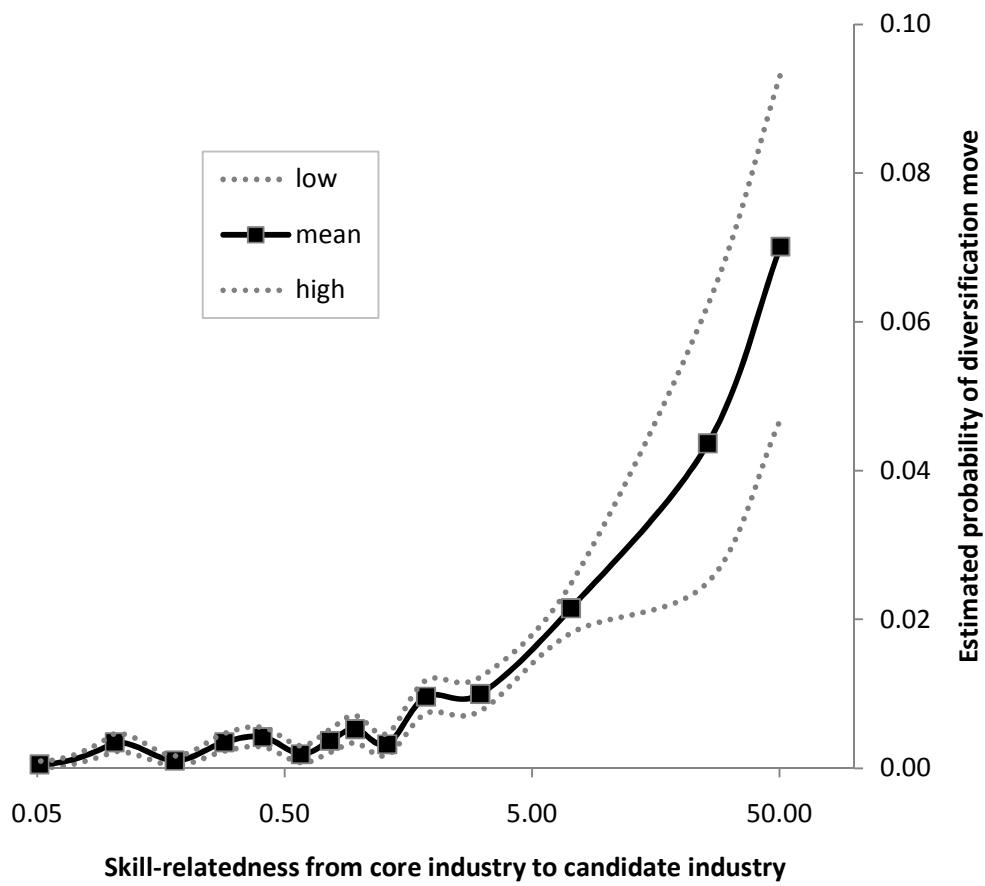


Figure 2. Skill-relatedness and estimated diversification probabilities.