Heterogeneity of Intellectual Assets – a Method for Identification and Measurement With Patent Data

Henrich Dahlgren*, Rasmus Lund Jensen and Finn Valentin

*hd.ivs@cbs.dk
Research Centre on Biotech Business, Copenhagen Business School
Solbjergvej 3,3, DK-2000 Frederiksberg (Denmark)

Abstract
This paper deals with methodological issues of measuring and assessing the composition and level of heterogeneity of firms' intellectual assets. It develops an original metric - referred to as the H-index - for measuring heterogeneity at firm level using data extracted from patent documents. The main purpose is to improve the characterization of research activities within firms in the biotechnology sector. Although the H-index grew out of research on biotech firms, the metric carries broader relevance for all patent-intensive industries. The measurement and assessment of the H-index is illustrated and tested using empirical data from our study on Scandinavian biotech firms.

Introduction

This paper addresses methodological issues of measuring and assessing the composition and level of heterogeneity of firms' intellectual assets. It presents an original metric for measuring heterogeneity, using data extracted from patent documents. We developed this metric - referred to as the *H-index* - with the purpose of improving the characterization of firms in the biotechnology sector. The core of this sector, Dedicated Biotech Firms (DBFs), has the production of (drug discovery) research as its sole economic activity, undertaken primarily on the basis of intellectual assets. This type of output is inadequately characterized by standard classifications of industries or products, rendering public statistics of limited use. In contrast, outputs from biopharmaceutical research are intensely protected by IPR, making patent documents an attractive alternative data-source, provided that patent data can be transformed into relevant indicators. Although the H-index grew out of research on biotech firms, the metric carries broader relevance for all patent-intensive industries. Terminologically speaking, the concept of "heterogeneity" in this context is preferred above its semantic cousin, "diversification". The latter emphasizes the extension of assets or activities from a given point of departure, and is often associated with an analytical focus on the strategy and direction of the change by which this extension comes about. In comparison, heterogeneity is a more straightforward phenomenon, essentially concerned with the degree of dissimilarity between constituent elements of a composite configuration, assuming no particular locus or level of original homogeneity.

First, the paper starts with a brief overview of the theoretical issues that have spurred an interest in measuring heterogeneity. Second, a brief review of the literature on diversification and heterogeneity reveals similarities and differences between various methods of measuring inter-firm and intra-firm heterogeneity, and we argue there is a need for an additional method. Third, we present and discuss the H-index in more detail. In this section, we describe the process of measuring intra-firm heterogeneity by transforming standard patents codes (IPC) into a corresponding classification system, referred to as *H-codes*. Finally, we discuss the assessment of the H-index and corresponding composition and level of heterogeneity and test the validity of the metric using empirical data from our study on Scandinavian biotech firms.

The theoretical context

A variety of theoretical issues in the literature on industrial dynamics have stimulated the development of different methodologies to measure heterogeneity and diversification. We present a brief review of how key theoretical issues have stimulated concerns regarding the measurement of heterogeneity as several of these methodologies appear, in modified form, in the metric developed in this paper.

Proceedings of ISSI 2005, pages 295–304.
Edited by P. Ingwersen and B. Larsen.

Diversification

Diversification and its effects on firm performance, has formed the basis for a key research question in the strategic management literature. Since performance differences, as an outcome of related and unrelated diversification, has been a major issue of concern in this research, a number of methodologies have been developed to measure the degree of heterogeneity between different parts of diversified firms (Chatterjee and Wernerfelt 1991; Ramanujam and Varadarajan 1989; Markides and Williamson 1994). A number of methodologies draw on the use of patent data (Breschi *et al.* 2003), and particularly their classification codes, related with the literature referred to below.

Managing technological diversity

Increasing diversification of the knowledge and technology base of companies accentuates the strategic issue of managing diversity. Theoretically, three key forces are argued to drive technological diversification within firms (Granstrand *et al.* 1997); (a) opportunities to improve products and production systems through the implementation of new technologies, (b) continuing competitive advantage of existing technologies, and finally (c) changes in products, production systems, and supply chains caused by innovation. Technological diversification is, in turn, argued to be a driving force behind, among others, firm growth, increasing R&D investment, and the emergence of new business opportunities based on related technologies (*ibid*.). Firms tend to manage more technologies to generate products and services than may be derived from observing the line of business they are engaged in (Granstrand 1998). Granstrand *et al.* (1997), for instance, found that firms tend to diversify into a higher number of technological fields than product classes, but the diversification tends to occur to a high extent within the realm of related technological fields (Patel and Pavitt 1997). From another perspective, complementarities may cause firms to diversify into different technological fields when, for instance, they wish to diversify their range of related products (Pavitt 1998). This is especially observed in large firms, although their technological profiles seem to be rather stable over time (Patel and Pavitt 1997). Diversification within firms may be related to learning spillovers (Henderson and Cockburn 1996). From this perspective, firms exploit knowledge externalities and economies of scope derived from the diversity of their activities (Henderson and Cockburn 1994).

Metrics

Diversification at the firm level has been measured with the use of data on input-output flows between industries. Scherer (1982) measured relatedness at industry level based on R&D expenses in one industry and the use of the generated output in other industries. The 'concentric index' developed by Caves *et al*. (1980) has been used in several subsequent studies, for example by Montgomery and Wernerfelt (1988), and bears some resemblance to the metric introduced in this paper. The concentric index measures diversification by taking into account the percentage of a firm's sales in different industries and a weighted value for the degree of similarity between SIC codes among the industries in which the firm is engaged. The index is calculated as

$$D_i = \sum_{j=1}^{n} m_{ij} \sum_{l=1}^{n} m_{il} r_{jl}$$

where D_i is the diversification of firm *i*, determined by the percentage of firm *i*'s sales in industry *j* (m_{ij}) and the relatedness of the industries in which firm *i* is engaged (r_{jl}). The latter factor is assigned value *0* if industry *i* and *l* have the same three-digit code, value *1* if industry *i* and *l* have the same two-digit code but different three-digit codes, and value *2* if they have different two-digit codes.

A number of studies have used *patent classification* data to indicate diversification. The International Patent Classification (IPC) system is used by patent examiners for categorizing and organizing patents within different technological fields[1]. Classification is based on the technological content and context of patents. Codes are dispersed among a variety of technological fields, primarily for search purposes, and have been increasingly used for measuring diversification among firms (e.g. Jaffe 1986; Jaffe 1989; Granstrand *et al*. 1997; Patel and Pavitt 1997; Patel and Pavitt 2000). Firms' technological

[1] For further information about IPC, please refer to www.wipo.int.

competencies and diversification into different technological fields are often measured using firms' patent shares in different technological fields or firms' revealed technology advantage. The former is simply the relative share of a firm's patent portfolio in a given technological field. The latter is defined as the relative importance of the firm in each field of technological competence, after normalizing for the firm's share of total patenting. Jaffe (1986) and Jaffe (1989) focused on the relatedness of technologies at the firm level. Jaffe measured technological relatedness between firms using patent classifications by applying the 'cosine index', which has been frequently used to measure the level of similarity of research in technological fields between firms. The similarity, or "technological proximity", is measured by calculating the overlap of the fractions of firm A's and firm B's patents in different patent classes. Technological proximity is, applying the cosine index, given by

$$P_{ij} = \frac{\sum_{k=1}^{K} f_{ik} f_{jk}}{\sqrt{\left(\sum_{k=1}^{K} f_{ik}^2\right)}\sqrt{\left(\sum_{k=1}^{K} f_{jk}^2\right)}}$$

where f_{ik} and f_{jk} are the fractions of firm *i*'s and firm *j*'s patents in patent class *k*. P_{ij} is the degree of overlap between f_i and f_j. P_{ij} is 1 when f_i and f_i are identical, and 0 for firms without any overlap of patents. Verspagen (1997) and Breschi *et al*. (2003) use the co-occurrence of patent classification codes to describe relatedness between different technological fields. The assumption is that strong relationships between technological fields, measured as the frequency of co-occurrence in patent classifications, may shed light on the degree of "knowledge relatedness" and knowledge spillover between technologies. Breschi *et al*. (2003) apply the cosine index to measure knowledge relatedness and its effect on firms' technological diversification. Firms' propensity to patent varies across industries so clearly there are sectors for which this type of data provides insufficient coverage. Limitations to the validity of patents have also been argued to stem from its inability to pick up the tacit dimensions of the knowledge of firms. Patel and Pavitt (1997) argue, however, that tacit and codified knowledge are complementary and, hence, reduces the limitations of patent data as a source for measuring heterogeneity.

Patent documents offer additional possibilities for characterizing the composition of intellectual assets. Citations to non-patent literature, primarily academic papers, are listed as document sources and antecedents of the inventions, which may be translated into information on the intellectual asset composition of firms (Narin 2000). Citations of previous patents may be applied in a similar manner (Granberg 1988). Patents also have text sections such as titles and abstracts allowing inventions to be characterized systematically. Such characterization may be extended to the entire patent portfolio of firms, which in turn offers indications of the composition of their knowledge and R&D assets. Characterization may be based on co-word analysis (e.g. van Raan and Engelsman 1993; Engelsman and van Raan 1994) and its further extension into text-mining methodologies (Valentin and Jensen 2003).

The H-index

The H-index is calculated for single firms, based on the main IPC codes of their patents. To build a metric particularly suited for the technological fields related to the biotech industry IPC codes are translated into an adjusted classification system, referred to as H-codes. Within specific technology fields, IPC codes offer categories so fine-grained that, for smaller quantities of patents, many categories are used too infrequently to accommodate statistical purposes. In these cases categories have been combined, based on their technological proximity.

In this translation into 3-level H-Codes the first level indicates the highest aggregation of technological fields, and level 3 the most detailed specification of technological fields. Based on an analysis of IPC codes assigned to the patents held by Danish and Swedish biotech firms, these codes

could be re-classified into nine different, homogenous categories at H-code level 1, and further categorized at level 2 and 3 in more detailed sub-categories. IPC code level 1 and 2 have been collapsed into H-code level 1, IPC code level 3 corresponds to H-code level 2, and IPC code level 4 to H-code level 3 (see further figure 1). Use of further levels of the IPC codes would generate too many categories at H-code level 4 and produce an excess level of detail for our study with a skewed number of patents in each subcategory[2].

The five IPC levels of IPC code "C12Q-001/18" denote the following:	
IPC-Level 1:	C: Chemistry and metallurgy
IPC-Level 2:	C12: Biochemistry; beer, spirits, wine or vinegar; microbiology or enzymology, mutation or genetic engineering.
IPC-Level 3:	C12Q: Measuring or testing processes involving enzymes or micro-organisms, compositions or test papers therefore, processes of preparing such compositions, condition-responsive control in microbiological or enzymological processes,
IPC-Level 4:	C12Q-001: Measuring or testing processes involving enzymes or micro-organisms and compositions therefore and/or processes of preparing such compositions.
IPC-Level 5:	C12Q-001/18: Measuring or testing processes involving viable micro-organisms testing for anti-microbial activity of a material.
This IPC code is translated into the H-code "7.2.0" in the following way: Category "7" corresponds to C12 in the IPC system. Category "2" corresponds to Q on IPC level 3. Category "0" corresponds to 001 on IPC level 4.	

Figure 1: Translation of IPC codes into H-codes

Calculating the H-index

The approach of the H-index is similar to Caves' concentric index in that it measures a weighted distance based on similarities between the codes representing different technological fields. An important difference compared to the ordered approach of Caves *et al*. (1980) is, however, the weighted and relative distance in our approach. The H-index is given by

$$H\text{-}index = \frac{\sum_{i=1}^{N-1}\left[\sum_{j=1}^{N-i} RV\left[P_i P_{i+j}\right]\right]}{\frac{N(N-1)}{2}}$$

The level of heterogeneity within each firm (*H-index*) is measured as the sum of the weighted values (*RV*) of patent H-code relationships $\left[P_i P_{i+j}\right]$ between all patents (*N*) held by a firm, normalized by the total number of relationships $\frac{N(N-1)}{2}$ between all patents. It is recorded whether H-codes are identical or different. Non-identical relationships are assigned specific *RV* for each of the three levels in the H-code. Differences at level 1 are assigned a *RV* of *1*. Differences at level 2 are given a *RV* of *0,5*, and differences at level 3 are assigned a *RV* of *0,25*. Relationships between patents with identical H-codes score a *RV* of *0*. Patent relationships $\left[P_i P_{i+j}\right]$ are assigned a *RV* on one level only. That is, if H-codes differ at a certain level between two patents, starting comparisons at level 1, no further evaluation is made at lower levels. Firms with none or one patent only are omitted from the calculation and not assigned any H-index value. Figure 2 exhibits the distribution of H-index on Danish and Swedish DBFs holding more than two patents each.

[2] For further information about the translation of IPC codes into H-codes, please contact corresponding author.

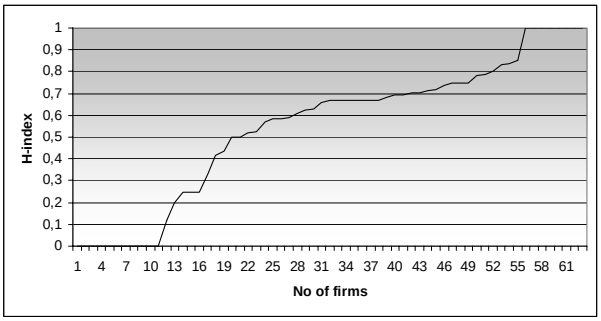

Figure 2: Distribution of H-index on Danish and Swedish DBFs (N=63)

To illustrate the calculation of the H-index and to see the effects of the number of patents and the distribution of patents in different technological fields, two cases of H-index calculation are shown in table 1. Patents of Firm A are found in three different technological fields at H-code level 1, indicated by three different codes (3, 5, and 8). Patents in category 3 at H-code level 1 are all found in category 2 at H-code level 2, and in category 3 at H-code level 3. Patents in category 5 at H-code level 1 are found in three different categories at H-code level 2, and so forth. To calculate the H-index, the H-code of patent 1 is compared with H-codes of patent number 2 to 10, the H-code of patent 2 is compared to the H-codes of patent 3 to 10, and so forth. For instance, the H-code of patent 1 compared to patent 2 and 3 respectively, scores a *RV* of 0 since they are identical. The H-code of patent 1 compared with patent 4 scores a *RV* of 1 since they differ at level 1. The H-code of patent 5 compared to that of patent 6 is assigned a *RV* of 0,5 since H-codes differ at level 2. The total *RV* is divided by the total number of relationships between patents, resulting in a H-index of 0,77. Firm B, on the other hand, hold patents in a single technological field. That is, the H-code is similar at all three levels. Consequently, all patent relationships are assigned *RV 0* because H-codes are identical and the H-index of Firm B is *0*. Hence, firm A exhibits higher heterogeneity than Firm B.

Table 1: H-indexes for a sample of two DBFs

FIRM A	H-index: 0,77	Patents: 10		
Patent no	**Main IPC**	**H-code**	**H-code**	**H-code**
1	A61K-031/045	3	2	3
2	A61K-031/167	3	2	3
3	A61K-031/167	3	2	3
4	C07C-275/00	5	1	5
5	C07D-207/00	5	2	0
6	C07K-014/705	5	5	3
7	G01N-033/48	8	0	2
8	G01N-033/50	8	0	2
9	G01N-033/50	8	0	2
10	G01N-033/68	8	0	2

FIRM B	H-index: 0,00	Patents: 4		
Patent no	**Main IPC**	**H-code**	**H-code**	**H-code**
1	A61K-031/12	3	2	3
2	A61K-031/135	3	2	3
3	A61K-031/136	3	2	3
4	A61K-031/5375	3	2	3

H-index and the composition of intellectual assets in biotech firms

To examine what the H-Index measures it is useful to distinguish between the four research approaches typically found among dedicated biotech firms: (1) bio-pharmaceuticals (mainly proteins

operating as drugs), (2) antibodies (belong to bio-pharmaceuticals but is analyzed separately because of substantial differences in targets and pathways), (3) small molecules (drugs based on molecules of lower complexity than the former two) or (4) delivery and diagnostics (focusing on methods by which drugs are delivered to patients and on tools for diagnosis).

These four approaches differ in the nature of their core problems, and hence also in the composition of the knowledge developed for their solution. A key issue in *bio-pharmaceuticals* (including antibodies) is to understand the complexity of lead molecules to an extent that permits them to be re-engineered so as to address highly specific targets and pathways for controlled therapeutic effects. In these research approaches, key intellectual assets must combine understandings of both complex pathways and intricate lead molecules. Their highly composite architectures expectedly should be reflected in high H-index values.

Small molecule drug discovery operates with leads of much lower complexity, but face the challenge of achieving complicated therapeutic effects by means of chemical design. For that reason, small molecule firms often focus on a specific binding site and the parts of therapeutic pathways that are immediately connected to that site, which in turn may relate to multiple disease groups. E.g. some receptor families are present in the membranes of many different cell-types in the human body. Correctly understood and approached, these receptors may open up to pathways relevant for multiple potential decease targets, and in small molecule approaches these potentials are more readily explored through the very high variability in synthetic compounds that may be generated and screened. For these reasons, small molecule firms build knowledge that is more focused on specific binding sites and their potential pathways. These efforts will typically be expressed in broader variability in the leads they explore and patent. Their knowledge heterogeneity therefore may be expected to be smaller and more directly associated with variability in target exploration.

Delivery and diagnostics, finally, represent highly focused research approaches, and we should expect them to have the lowest knowledge heterogeneity among the four approaches.

Differences between the four discovery approaches are presented in figure 3 which gives mean and inter-quartiles of the H-index in each research strategy, confirming that bio-pharmaceuticals and antibodies show higher mean H-index and their distributions are concentrated at higher levels of H-index.

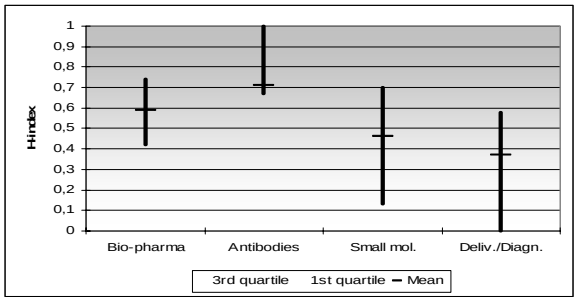

Figure 3: Distribution of H-index on research approaches indicated by the 1st and 3rd quartile in combination with the mean value (vertical line)

To understand these differences we need a few general concepts on the types of knowledge assets that DBFs bring to bear on their discovery processes. The range of competencies in DBFs involved in drug discovery may be classified in three main types of intellectual assets (Valentin *et al*. 2005):

- *Conceptual frameworks* include theories, models and heuristics specifying or suggesting causal relationships and the conditions under which they are operative.
- *Methods* include tools, procedures and research instrumentation for generating, processing and interpreting data. High-throughput screening is probably the most well-known data processing tool for effective screening of high amounts of data

- *Internally generated information* such as screening libraries or other results of previous transformations of data into higher-order inputs for problem solving.

The differences between the four approaches summarised above also means that they vary substantially in the ways they draw on and combine the three asset types, and in turn that should produce different values on the H-Index. To examine differences in the way these knowledge assets are combined in discovery activities we apply a simplified categorization of patents based on their relationships to the three knowledge assets, as identified by their main IPC. Patents referring directly to compounds and to protein leads, we argue, relate particularly to the conceptual framework assets of DBF, because they are embodied expressions of the way targets and pathways are perceived and modelled. These compound patents will be referred to as COP. The remaining patents of firms, not relating to specific compounds or leads, will instead represent the other two types of assets, i.e. methods and information, and they are referred to as tools/method patents (TMP). The share of COP and TMP patents in firms pursuing each of the four research approaches is presented in figure 4.

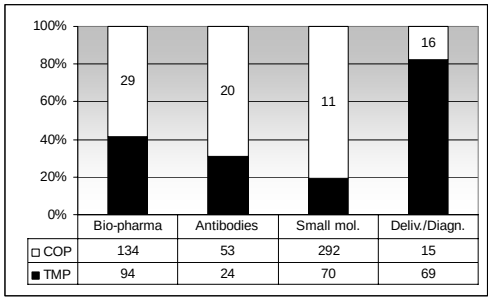

Figure 4: Share TMP and COP patents in research approaches
(number of firms indicated inside and number of patents beneath columns)

Antibodies and other bio-pharmaceuticals are seen in figure 4 to tend towards more equal shares of COP and TMP patents, whereas small molecule firms and firms in diagnostics are more one-sided, having predominantly one type of patent. The two latter research approaches, accordingly, have *less composite* knowledge structures than do the two former.

Assessment of H-index and final remarks

To assess the impact of COP and TMP patents on the H-index, we decompose the overall *RV* for each firm into three separate components; a *RV* associated with H-code differences *within* COP patents (RVCOP), a *RV* associated with H-code differences *within* TMP patents (RVTMP), and a *RV* associated with H-code differences *between* COP and TMP patents (RVMIX). The sum of these three components corresponds to the total *RV* as obtained in the H-index calculation. Each sum is divided by the aggregate sum, revealing their share of the total *RV*. For each firm the share of total *RV* for RVCOP is calculated, along with corresponding shares for RVTMP and RVMIX. The average shares for firms in each research approach are presented in figure 5.

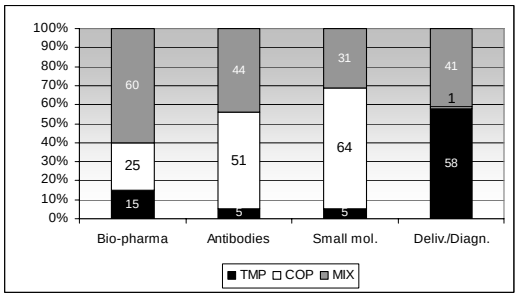

Figure 5: Share of total *RV* for patent categories in each research approach.

The total *RV* for biopharmaceutical firms has a very high share of RVMIX (60%), while the second highest share (44%) is observed for antibody firms. Small molecule firms in particular reveal high impact of RVCOP (64%). Delivery and diagnostics exhibit a high impact of RVTMP (58%) and of

RVMIX (41%). Figure 5 confirms that each of the components of RVCOP, RVTMP AND RVMIX has a share of overall *RV* corresponding to what we should expect from the composition of COP and TMP patents presented in figure 4. The one exception is that delivery and diagnostics in figure 5 displays a high share of RVMIX, although figure 4 showed them as having predominantly TMPs as their only type of patents. This distortion emerges as an effect of isolating, at this specific stage of the argument, each of the components, which is not the case in the full version of the H-index[3].

We test the impact of heterogeneity within each patent category on the H-index, by calculating separate *sub-indexes* for heterogeneity for TMP patents (HidxTMP), COP patents (HidxCOP), and for dissimilar patent relationships (HidxMIX). These sub-indexes are calculated in the same way as the main H-index as per 2003 (Hidx03), where the total *RV* in each category is divided by the total number of patent relationships in each category. These sub-indexes are regressed on the H-index (Hidx03) to determine its sensitivity to each sub-index respectively. We test relationships with each of the sub-indexes separately, despite few observations in each research approach, because (a) together they explain the full variation of the H-index and (b) when regressed in pairs they cause too high levels of autocorrelation.

Table 2 exhibits the results for HidxTMP. The findings exhibited in table 2 indicate a relationship between heterogeneity within TMP patents and the H-index (Hidx03) for delivery and diagnostics and antibodies. That is, the level of heterogeneity in delivery and diagnostics and antibodies is driven by heterogeneity between TMP patents. This patent category has also the largest share of patents in delivery and diagnostics.

Table 2: Regression analysis for HidxTMP

Independent variables		***Dependent variable Hidx03***			
		Model 1: Bio-pharma firms (N = 14)	Model 2: Small molecule firms (N = 7)	Model 3: Delivery or Diagnostic firms (N = 8)	Model 4: Antibody firms (N = 4)
Intercept		0.5288*** (0.0881)	0.3545* (0.1576)	0.1085 (0.0889)	0.6535*** (0.0328)
HidxTMP		0.1742 (0.1664)	0.5769 (0.2875)	0.9094*** (0.1902)	0.2176* (0.0546)
Model	**Pr > F**	0.3159	0.1011	0.0031	0.0576
	Adj R-sqr	0.0073	0.3352	0.7576	0.8322
	Df	13	6	7	3
Standard errors are given in parentheses for each estimate. Levels of significance as indicated by * = 10% level, ** = 5% level and *** = 1% level.					

Table 3 exhibits the results for HidxCOP. The results indicate a relationship between heterogeneity within COP patents and the H-index (Hidx03) for antibodies, small molecules, and bio-pharmaceuticals. Their level of heterogeneity seems to be driven by heterogeneity between COP patents. This patent category exhibits the largest share of patents in small molecules and seems to exhibit higher impact on the H-index than TMP patents in antibodies and bio-pharmaceuticals.

[3] This distortion emerges because the 15 COP patents are dispersed with most of 16 firms in this approach having only one COP. That is, variation within COP patents disappears, whereas their dissimilarity from TMP patents in the same firms appears as high RVMIX shares.

Table 3: Regression analysis for HidxCOP

Independent variables		***Dependent variable Hidx03***			
		Model 1: Bio-pharma firms (N = 18)	Model 2: Small molecule firms (N = 15)	Model 3: Delivery or Diagnostic firms (N = 3)	Model 4: Antibody firms (N = 11)
Intercept		0.4299*** (0.0899)	0.0074 (0.0487)	0.0206 (0.0564)	0.2145* (0.1158)
HidxCOP		0.4263* (0.2151)	1.1763*** (0.1033)	1.4555* (1.9512)	0.8188*** (0.1772)
Model	**Pr > F**	0.0649	0.0001	0.0848	0.0013
	Adj R-sqr	0.1469	0.9019	0.9647	0.6705
	Df	17	14	2	10
Standard errors are given in parentheses for each estimate. Levels of significance as indicated by * = 10% level, ** = 5% level and *** = 1% level.					

Table 4 shows the results for HidxMIX. The findings indicate a relationship between heterogeneity within combination of patents and the H-index (Hidx03) for bio-pharmaceuticals. The level of heterogeneity in bio-pharmaceuticals appears to be driven by heterogeneity in combinations of patents. Tools/method patents and compound patents have similar shares of patents in bio-pharmaceuticals.

Table 4: Regression analysis for HidxMIX

Independent variables		***Dependent variable Hidx03***			
		Model 1: Bio-pharma firms (N = 18)	Model 2: Small molecule firms (N = 8)	Model 3: Delivery or Diagnostic firms (N = 5)	Model 4:*) Antibody firms (N = 7)
Intercept		0.1183 (0.1376)	5.8514 (3.7559)	0.1562 (0.1613)	-
HidxMIX		0.6151*** (0.1468)	-5.1593 (3.7646)	0.4804 (0.2270)	-
Model	**Pr > F**	0.0007	0.2196	0.1246	-
	Adj R-sqr	0.4934	0.1115	0.4651	-
	Df	17	7	4	0
Standard errors are given in parentheses for each estimate. Levels of significance as indicated by * = 10% level, ** = 5% level and *** = 1% level. *) model is full rank since the degree of freedom is equal to zero					

The findings reveal an impact on the level of H-index primarily of heterogeneity within patent categories that relates to important intellectual assets and combination of those assets in different research approaches. The construct of the H-index enables analysis of the extent to which different types of intellectual assets drive the level of the H-index and to what extent the heterogeneity within different assets impact the overall level of the H-index depending on the size of their share of patents. Hence, the metric allows analysis of the strategic importance of firms' intellectual assets and the combination of those assets.

The H-index may be used for analysis of firms in all patent-intensive sectors. In a parallel paper the authors demonstrate its usefulness for unpacking and analysing architectures of knowledge assets and their related scope advantages in biotech firms (Valentin *et al*. 2005).

References

Breschi, S., F. Lissoni and F. Malerba (2003) 'Knowledge-relatedness in firm technological diversification', Research Policy, 32(1), 69-87.

Caves, R. E., M. E. Porter and A. M. Spence (1980) 'Competition in the Open Economy: A Model Applied to Canada', Cambridge, MA.: Harvard University Press.

Chatterjee, S. and B. Wernerfelt (1991) 'The Link between Resources and Type of Diversification: Theory and Evidence', 12, 33-48.

Engelsman, E. C. and A. F. J. van Raan (1994) 'A Patent-Based Cartography of Technology', Research Policy, 23, 1-26.

Granberg, A. (1988) 'Fiber Optics as a Technological Field', Research Policy Institute, Lund University.

Granstrand, O. (1998) 'Towards a theory of the technology-based firm', Research Policy, 27, 465-489.

Granstrand, O., P. Patel and K. Pavitt (1997) 'Multi-Technology Corporations: Why They have "Distributed" Rather than "Distinctive Core" Competences', California Management Review, 39(4), 8-25.

Henderson, R. M. and I. Cockburn (1994) 'Measuring Competence? Exploring Firm Effects in Phamaceutical Research', Strategic Management Journal, 15, 63-84.

Jaffe, A. B. (1986) 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value', American Economic Review, 76(3), 984-1001.

Jaffe, A. B. (1989) 'Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers', Research Policy, 18(2), 87-97.

Markides, C. C. and P. J. Williamson (1994) 'Related Diversification, Core Competencies and Corporate Performance', Strategic Management Journal, 15, 149-166.

Montgomery, C. A. and B. Wernerfelt (1988) 'Diversification, Ricardian Rents, and Tobin's q', RAND Journal of Economics, 19(4), 623-632.

Narin, F. (2000), 'Assessing Technological Competencies', in J. Tidd (ed.), 'From knowledge management to strategic competence. Measuring technological, market and organisational innovation', London: Imperial College Press, 155-196.

Patel, P. and K. Pavitt (1997) 'The technological competencies of the world's largest firms: complex and path-dependent, but not much variety', Research Policy, 26, 141-156.

Patel, P. and K. Pavitt (2000), 'How Technological Competencies Help Define the Core (not the Boundaries) of the firm', in G. Dosi, R.R. Nelson and S.G. Winter (eds), 'The nature and dynamics of organizational capabilities', Oxford: Oxford University Press, 311-333.

Pavitt, K. (1998) 'Technologies, Products and Organization in the Innovating Firm: What Adam Smith Tells us and Joseph Schumpeter Doesn't', Industrial and Corporate Change, 7(3), 433-452.

Ramanujam, V. and P. Varadarajan (1989) 'Research on corporate diversification: A synthesis', Strategic Management Journal, 10, 523-551.

Scherer, F. M. (1982) 'Interindustry Technology Flows in the United States', Research Policy, 11, 227-245.

Valentin, F. and R. L. Jensen (2003) 'Discontinuities and distributed innovation - The case of biotechnology in food processing', Industry and Innovation, 10(3), 275-310.

Valentin, F., H. Dahlgren and R.L. Jensen (2005) 'Research Strategies in Small Science-based Firms: Effects on Performance in Danish Biotechnology', Paper presented at Druid Summer Conference, Copenhagen, June 26-30

van Raan, A. F. J. and E. C. Engelsman (1993) 'International Comparison of Technological Activities and Specializations: A Patent-based Monitoring System', Technology Analysis & Strategic Management, 5(2), 113-137.

Verspagen, B. (1997) 'Measuring Inter-sectoral Technology Spillovers: Estimates from the European and US Patent Office Databases', Economic Systems Research, 9(1), 49-67.