INDUSTRY CLASSIFICATIONS: AIM, SCOPE AND TECHNIQUES*

MICHAEL PENEDER

AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH (WIFO)

Corresponding author:

Michael Peneder
Austrian Institute of Economic Research
(WIFO)
P.O. Box 91; A-1030 Vienna
Austria
Tel.: + 43-1-798 26 01 – 480
Fax.: + 43-1-798 93 86
E-mail: michael.peneder@wifo.ac.at

* The author gratefully acknowledges the support of the European Commission, under the fifth framework project number HPSE-CT-2001

Abstract

Industry classifications select essential characteristics of technology and markets, condensating the vast heterogeneity of competitive environments into a smaller number of salient types. Although frequently applied in empirical studies on industrial economics, technological development, international trade, and competitiveness, we still find little or no methodological discussion and a striking lack of awareness for the different approaches pursued. This interpretative survey is a first step in the collection of systematic information about the aim, scope and techniques relevant to the major classifications currently used in applied economic studies.

Key words: structural analysis, taxonomies, technological regimes, factor intensities, intangible investment

JEL Codes: B41, C81, F14, L60, L80

1. Introduction

Classifications systematically arrange cases in terms of their similarity. They constitute a first and generic advance from mere observation and description towards systematic scientific inquiry. Substituting structural knowledge for exhaustive information about single attributes, the intractable diversity of real-life phenomena is condensed into a smaller number of salient types. Classifications thus direct our attention towards a few characteristic dimensions, according to which relative similarities or differences can be identified. They allow us to take account of heterogeneity, but simultaneously force us to be selective.

In the field of economics, a historical example of such a classification is Carl Menger’s (1871, p. 21f) distinction between goods of the “first order” and goods of “second”- and “higher orders”. While the former refers to consumer goods, which directly serve our wants and needs, the latter categories are comprised of intermediate goods, such as capital, labour or raw materials, which are used to produce goods of the first or the respective lower order. This structural separation was an important element in the formulation of Menger‘s theory of economic value and prices. The related separation of industries according to major *end-use categories*, such as “intermediate goods”, “equipment”, or “consumer goods” has remained popular up to the present day.

Apart from this example, one can argue that more generally, the structuralist perspective has strong intellectual roots in the Austrian School of economics. Streissler (1969) even claims that it was the structural mode of thought which above all was handed down by Carl Menger to his successors and which eventually became their most characteristic common denominator. More than their association with the great inventions of subjective value theory and marginal utility at the beginning of the neo-classical revolution, Streissler claims that the structural approach, its “focus on heterogeneity as a central theoretical vision” (ibid. p. 253), truly singled out and unified the Austrian School. Structural analysis is thereby defined as "the decomposition of aggregates in order to increase the explanatory content which can be derived by viewing the aggregates as undifferentiated wholes" (ibid. p. 241). This definition of structural analysis suits our purpose particularly well, because it emphasises the analytic support of the generation of knowledge by means of structural separation. Similarly, in our survey, we will limit our considerations to industry classifications which support structural analysis of the type described above; that is, those which are constructed with a specific analytical purpose.[1]

The process of classification is generally defined as the ordering of cases in terms of their similarity. According to Bailey (1994, p. 5ff), classifications themselves can be distinguished by (among others) the following characteristics: They can be labelled either as typologies or taxonomies; monothetic or polythetic; synchronic or diachronic. The term *typology* refers specifically to a conceptual classification, the cells of which represent type concepts rather than empirical cases. Conversely, the term *taxonomy* refers to a classification of empirical entities based upon quantitative analysis. In this sense, one can also distinguish *monothetic* classes, in which all the cases included in a certain category are identical with respect to every relevant dimension. No exceptions or further differentiations are allowed. Such a neat and (idealised) categorisation is typical of qualitative categorisations, whereas empirical classifications generally come up with *polythetic* classes. Here, the cases are not identical with respect to all variables, but rather are grouped according to the generally strongest similarity. Finally, classifications are called *synchronic* (or phenetic), if they refer to the characteristics of an observation at a certain point in time. Conversely, classifications are called *diachronic* (or phyletic), if they are based upon characteristic patterns of change or evolution. All the current classifications discovered for this survey are synchronic in the sense that they are static snapshots of structural characteristics, without dynamic categories for different patterns of change. Furthermore, the classifications are expected to be *exhaustive* and *mutually* *exclusive*, thereby demanding the existence of one (but only one) appropriate class for each observation. In practice, this can pose problems, especially if it is nearly impossible to discriminate between many observations or conversely, if many outliers exist. In the first case, the boundaries between groups are difficult to draw, while in the second case, too many classes arise.

Two general approaches to the quantitative identification of individual observations into classes can be distinguished. A 'cut-off' procedure by which a certain discriminatory edge is defined exogenously by the researcher is the more frequently applied method. The sole advantage of this approach lies in its simplicity. In choosing not to use more powerful statistical tools, the underlying structure within the data is more or less presumed, rather than explored. Although this approach can be defended as long as the classifications are built upon one or two variables only, it is generally inept for the categorisation of a data profile of larger dimensions. Statistical cluster analysis is the obvious alternative. It is a powerful technique specifically designed for classifying observations on behalf of their relative similarities with respect to a multidimensional array of variables. The basic idea is one of dividing a specific data profile into segments by creating maximum homogeneity within and maximum distance between groups. Despite its higher technical sophistication, cluster analysis is still a heuristic method, which requires the researcher to make a number of critical choices, e.g. between the various measures of distance or the kind of agglomeration algorithm. The variety of possible outcomes naturally raises concerns about the robustness of the results and calls for an extensive documentation of the work done (including information about the approaches which were dismissed). Cluster analysis also requires a clear concept of geometric space that allows for a meaningful measurement of distances between observations. Variables have to be chosen in a way that spans independent dimensions of the phenomenon under investigation.

Today, the structuralist perspective appears most frequently in empirical studies on competitive performance, technological development, international trade, and industrial economics. But in contrast to the prominent attention it is given in various sciences such as biology, psychology or sociology, the proper construction and use of classifications has remained highly under-researched within the realm of economics. We still find little or no methodological debate and a striking lack of awareness for the different approaches pursued.[2] Although this survey tries to open up the discussion by collecting systematic information about the aim, scope and techniques involved in the major classifications currently used, our main objective is to provide the practitioner in the field with a brief overview for better orientation. Our focus will generally be on taxonomies, i.e. empirical classifications generated by quantitative identification. Only brief references will be made to influential typologies, which are purely based on conceptual considerations. Obviously, coverage can not be comprehensive.

In contrast to the summary in Table 1, which orders taxonomies chronologically, the main body of the text is organised around the major topics pursued by the various classifications. In Section 2, we begin with the one-dimensional separation of high- versus low-tech industries,[3] which frequently appears in studies on competitive performance by international organisations and governmental bodies. The classifications presented in Section 3 are motivated by the more differentiated notion of technological regimes and are developed within the field of innovation research. Section 4 discusses a distinct kind of classification, based upon various combinations of different factor intensities and closely associated with studies on international trade. Section 5 extends the perspective from conventional tangible factor intensities to the more recent emphasis on intangible investment (and endogenous sunk cost) in industrial economics. An entirely different approach, which targets the distinction between ‘quality‘- versus ‘price competition‘, is briefly discussed in Section 6. Section 7 presents the summary table and concludes.

2. The big divide: high-tech versus low-tech

The catchy distinction between high- and low-tech industries certainly takes the prize for the most popular industry classification. Although easily accessible in general terms (everyone has at least some idea of what it means), its conceptual foundations are fuzzy. Due to the lack of a precise definition, no clear and commonly acknowledged criterion for selection has been established. As a consequence, various governmental institutions, research organisations and other entities constructed a different classification, following somewhat divergent ideas about the meaning of ‘high-tech‘, which were hardly comparable. Most of them were based on certain quantitative measures of R&D expenditures, R&D personnel, or patent statistics. This variety of approaches expanded further, when at the beginning of the 1990s, a new kind of classification emerged, based upon subjective expert opinion.

The two representative examples of subjective identification are the classification of ‘advanced technology products‘ (ATPs) established by the U.S. Census Bureau and presented in McGuckin et al. (1992) and a Eurostat list of high-tech products documented in Capelot and Lambertz (1993). The methodological procedure is in both cases almost identical: “As a first step, a number of technological areas were identified in which ‘rapid developments of new concepts and ideas have lead to many breakthroughs that could be considered on the leading edge of modern technology‘“ (Capelot and Lambertz, 1993, p. 12). In a second step, individual products were identified as ATPs or high-tech products, if individual industry analysts determined that they contained significant amounts of one or more of these technologies. In addition, Capelot and Lambertz reported that a number of pragmatic aspects, such as the “complexity of the production process“, the “newness“ of a product, or the "unintelligibility of the function of a product (the less the function of a product is understood by a normal person, the higher the tech)" (ibid.) were taken into account.

Lifting the limitations which arise in connection with the availability of detailed industry data on R&D intensity, etc., the major advantage of subjective identification lies in the higher degree of disaggregation which can thus be achieved. Its advantage is therefore best exploited when it is applied to products in trade statistics, which generally are more disaggregated than industry data. Besides the enormous effort necessary for the installation and organisation of a sufficiently high-powered panel of industry experts for repeated revisions, the major shortcoming of this approach lies in the unavoidable distortions caused by the choice of experts. According to what criteria should they be selected? Obviously, different panels also imply different classifications. Since the decision process is not at all transparent, it is almost impossible to compare them in terms of quality or adequacy. One must conclude that, despite its obvious merits, subjective identification has further increased the general confusion about the multitude of available classifications.

For the purpose of international comparative studies on competitive performance, convergence on a general and methodologically reliable standard classification would however be desirable. The following high-tech classifications produced by the OECD might be considered the best choice. Trying to combine the benefits of quantitative and subjective methods, Hatzichrinoglou (1997) presents a sector- and a product-based classification. The sector classification is broad, targeting 22 ISIC 2-digit manufacturing sectors and relies exclusively on quantitative data for R&D intensity. In addition to R&D expenditures, an indirect measure of R&D embodied in intermediate and capital goods is also calculated on the basis of input-output matrices. The data employed are from the ANBERD database and cover the years 1980 to 1990 for selected OECD countries. A sector's total R&D intensity is the sum of direct and indirect intensities. The final classification is therefore one-dimensional and comprised of the following four groups: high technology; medium-high, medium-low, and low technology.

The product classification combines a quantitative cut-off procedure (i.e. the simple separation of observations by a given threshold value) at higher levels of aggregation, with a subjective evaluation for more disaggregated products. In 1994, an initial list was prepared in conjunction with the German Frauenhofer Institut (ISI) at the level of SITC 3-digit product groups. Groups were selected, which were characterised by a high R&D intensity (R&D to sales ratio) in a sample of six OECD countries. This selection was then further modified by means of subjective identification for the 5-digit products. The purpose was to make it largely compatible with the sectoral list. Although the identification is in general more restrictive, a number of high-tech products were also shown to originate in the medium-high technology category of the sectoral classification.

Finally, covering manufacturing and the services sector at a very broad level of aggregation, the OECD has more recently relied on the notion of 'knowledge-based' industries. By and large, this concept reflects the same mode of thought as the older high-tech versus low-tech terminology; however, it is more suitable for the inclusion of service industries. The OECD considers industries to be knowledge-based, when they are "intensive users of high technology and/or have a relatively highly skilled workforce that is required to benefit fully from technological activities" (OECD, 1999, p. 18]. This definition enables us to distinguish industries according to the relative importance of average research expenditures and the shares of skilled labour in total employment.[4]

3. Technological regimes

Compared to the simple high-tech versus low-tech distinctions, a more ambitious and refined concept had already emerged in the field of innovation research by the beginning of the

1980s. Deeply moulded into the evolutionary tradition, it stresses the diverse and contingent nature of competitive behaviour. Within the confines of this paradigm, competitive performance depends on the capability to match a firm's organisation and strategy to the technological, social and economic restrictions imposed by its business environment. It is in particular the notion of *technological regimes* (Nelson and Winter, 1982; Winter, 1984; Malerba and Orsenigo, 1993) or relatedly, of technological *paradigms* (Dosi, 1982, 1988), which puts especial emphasis on the importance of sectoral patterns in technological change. According to Malerba and Orsenigo (1993, p. 46) a technological regime describes the technological environment in which firms operate, characterised in terms of opportunity, appropriability, cumulativeness, and the complexity of the knowledge base. In his simulation model, Winter (1984) applied a rudimentary conceptual distinction between 'entrepreneurial' and 'routinised' regimes, referring to two distinct modes of Schumpeterian innovation.[5]

Combining quantitative information with visual inspection and inductive reasoning, Pavitt (1984) created an empirical classification of sectors according to the characteristic technological trajectories of its innovating firms. The data base is comprised of 2000 innovations considered as being 'significant', and the corresponding innovating firms in the UK; the time span ranges from 1945 to 1979. The data on innovations cover three- and four-digit product groups, which accounted for about half of British manufacturing output. Since the data are arguably very complex and incomplete, Pavitt refrained from a purely deductive approach with advanced statistical techniques. Instead, he aggregated the information into 11 groups at the 2-digit level and 26 groups at the levels of 3- or 4-digit products. He explains that this "approach has the advantage of allowing the patterns of the statistical data to be compared to the mind's eye with the rich range of sectoral and firm studies of technical change that have accumulated over the past 25 years" (ibid., p. 345). His attribute space is comprised of three major dimensions along which sectors are compared: (a) the 'sectoral sources of technology used‘, particularly in reference to the relative importance of purchases within or external to the sector; (b) the ‘institutional sources and nature of the technology produced in a sector‘, distinguishing, for example, whether the case in question deals with product or process innovations, or whether there are internal sources of knowledge, inputs from other firms or from public infrastructure; and finally (c) the ‘characteristics of innovating firms‘, i.e. their size as well as their main activity.

Pavitt's taxonomy of ‘sectoral technological trajectories‘ classifies industries as characterised either by (i) ‘science based‘ firms; (ii) ‘production intensive‘ firms, or (iii) ‘supplier dominated‘ firms. The second group is further subdivided into the categories of ‘scale intensive‘ production or ‘specialised suppliers‘. In the science based industries, the firm's own R&D expenditures are particularly high, although knowledge flows from universities and other institutions are also substantial. Due to the high R&D intensity, this group largely overlaps with the common identification of high-tech sectors. The technological trajectories of scale intensive producers in the production intensive category are characterised by high levels of in-house R&D targeted at process innovation and cost cutting. Firms are typically of larger size. This is in contrast to the comparatively small- and medium-sized production intensive suppliers, who focus primarily on product innovations for use in other sectors. Finally, supplier dominated firms have only small capacities for in-house R&D and engineering skills.[6] Apart from traditional manufacturing, most of the services sector is also located within this category.

Pavitt's taxonomy has been extremely influential, shaping the basic conceptual categories for the series of related classifications which followed. In the 1990s the availability of firm data from national innovation surveys induced a series of papers which are strongly related to the tradition of the Pavitt classification, but more critical of the presumed sectoral regularities in innovation patterns. Rather than classifying industries or sectors, they focus on the distinct innovation types observed at the firm level. Although the lack of industry classifications puts them outside our immediate concern in this survey, we should briefly mention *Cesaratto and Mangano* (1993), who applied an extended multivariate analysis (principle component and cluster techniques) to firm level data from the Italian survey on technological innovation in 1988. They concluded that their "results strengthen Pavitt's suggestion as to the fertility of classification of variety of technological patterns showing a similar, although more complex, combination of technological input and output. Moreover, while Pavitt links his taxonomy to representative sectors, our results lead us to emphasize individual firm strategy" (ibid., p. 252). Similarly, in two recent papers by *Arvanitis and Hollenstein* (1998) on manufacturing firms and *Hollenstein* (2001) on service firms, statistical cluster analysis and principle component techniques are applied to data from the Swiss Innovation Survey. Again, the authors primarily emphasise the observed variety of technological behaviour within sectors. Similarly, Archibugi (2001) argues for leaving aside Pavitt's link to industries and focus instead on the direct classification of firms.

It turns out to be one of the major virtues of the systematic collection of micro data from the innovation surveys, that the great heterogeneity of behavioural patterns at the micro-level can no longer be denied. In other words, assuming (as neoclassical economists do) that the competitive environment determines corporate strategy in its entirety (to the extent that firms are either forced to adopt optimal practices or to exit), implies an unwarranted denial of heterogeneity in competitive behaviour. However, any approach which neglects market specifics, conveys an equally unwarranted denial of heterogeneity in the firm's competitive environment. At this point, we also wish to add another practical remark. All of the industry classifications presented in this survey were initially aimed at serving a very specific analytic purpose. They were created because industry classifications enable the transposition of partly hidden characteristics, which are otherwise not readily available in internationally comparable formats, and then combine them with other, more easily comparable data (as long as these can be identified by industry-membership). It is difficult to find a similar, practical use for classifications which are only tied to individual patterns of behaviour. As a convenient means of condensing information, it is of course important for explanatory purposes. But since all analytical applications are locked into the same specific data set, one might generally prefer to work with the initial micro-data, where information is fully preserved.

Returning to our concern for industry classifications in the tradition of Pavitt's taxonomy, *Marsili* (2001) has recently offered a detailed and updated empirical account of the various 'modal characteristics of innovative processes' in relation to sectoral systems within manufacturing. Her methodology is similar to Pavitt's, in that it attempts to achieve a systematic summary of the available quantitative evidence from the rich SPRU data sources, thereby blending statistical analysis with inductive methods of systematisation and synthesis.[7] Her final classification is comprised of five groups: (i) the 'science based regime' with the highest level of technological opportunity (especially for product innovation) and strong links to academic research; (ii) the 'fundamental process regime', characterised by medium levels of opportunity (mostly for process innovation) and substantial benefits from academic research; (iii) the 'complex systems', distinctly characterised by the large differentiation of technological competencies and mostly indirect contributions from academic research; (iv) the 'product engineering regime' with a medium-high level of technological opportunities but a low persistence of product innovation; and finally, (v) the 'continuous-processes regime', characterised by low opportunities and a low persistence of innovations, while benefiting from upstream sources of embodied technological progress through the purchase of capital goods.

Finally, *Evangelista* (2000) presents an empirical Pavitt-type sectoral classification of innovation patterns in services.[8] Using firm level data from the Italian innovation survey for the period 1993 - 1995, eleven variables (ranging from total expenditures on innovation per employee, the relative importance of process versus product innovation, the shares of innovation costs devoted to R&D, design, or capital investments, to the dominant patterns of interaction between users, competitors and suppliers) were aggregated by factor analysis. Prior to the analysis, the data were reduced by aggregation into 22 broad sectors. In the end, Evangelista's classification comprised four main types of service industries: (i) the 'technology users', which resemble Pavitt's category of supplier dominated trajectories, where innovation costs are largely due to the purchase of capital equipment; (ii) the 'science & technology based' services, which are "located very upstream in the knowledge-generating chain" (ibid., p. 212), and which find their corresponding category in Pavitt's classification for the manufacturing sector; (iii) 'interactive- and IT-based' services, which are characterised by large investments in software and close interaction with the users; and (iv) 'technical consultancy', which is grouped separately because its characterisation largely overlaps with the second and the third categories.

4. Focus on factor intensities

In contrast to the differentiated patterns of technological regimes in the above innovation studies, international trade analysis emphasises factor intensities such as capital, labour, or natural resources, so that the patterns of specialisation can be tied to a location's relative factor endowments. For example, *Lawrence* (1984, p. 63ff) refers to a common conceptual separation between (i) 'Ricardo goods', which are particularly resource-intensive; and (ii) 'Hecksher-Ohlin goods', which use relatively standardised production technologies and are further subdivided into industries which are either particularly capital- or labour-intensive. Like many others, he then also adds a group of (iii) 'high technology goods', characterised by high proportions of R&D or scientists in the respective workforce.[9]

Legler (1982) and Schulmeister (1990) offer a remarkable empirical classification based on factor intensities, taking into account a broad set of factors ranging from the traditional intensities of capital, labour and raw materials up to human skills and expenditures on research and development. The initial data are for Germany and were produced by *Legler* (1982). He used a broad set of variables in order to characterise the typical production process within an industry: human capital, research and development, the investment of physical capital, labour inputs, economies of scale, inputs of raw materials and energy, as well as investments necessitated by environmental concerns. The amount of information is substantially reduced by a simple cut-off procedure.[10] After converting from industry to trade statistics, he determined an initial data matrix of ten qualitative binary variables (‘yes‘ or ‘no‘) times 154 manufacturing products (SITC 3-digits). The allocation of product groups is comprehensive but not mutually exclusive. Legler’s list of industries therefore does not yet qualify as a true classification in the sense we are seeking. It nevertheless gives an interesting collection of dummy variables based on a very broad set of factor intensities.

Aiming to map the ‘technology profile‘ of individual countries based on their patterns of trade specialisation, *Schulmeister* (1990) then undertook the complex task of recombining all the product groups into an exhaustive and mutually exclusive classification. In a very careful investigation he truly tested the limits of ‘visual‘ inspection for multidimensional data. The resulting taxonomy is a finely-grained hierarchy comprised of four layers, which split 154 product groups into the five broad groups of high-skill-, capital-intensive, labour-intensive, and resource-intensive technologies plus a residual category of other goods. The first group of human capital-intensive production is further separated into high-tech, low-tech, and others, which are divided further into labour- and capital-intensive goods. An additional level separates goods which are high-skill, medium-tech, capital-intensive and resource-intensive.

Similarly, the broad class of resource-intensive goods is subdivided into ‘strong‘ and ‘moderate‘ resource dependent products.

Schulmeister’s taxonomy stands out due to its high level of complexity, which on the one hand can be attributed to the many dimensions taken into consideration, and on the other hand has to do with his hierarchical solution to the problem of mutually exclusive identification. His taxonomy is certainly the most finely-grained classification surveyed in this paper. But its complexity might also be considered a weakness, producing too many classes containing too few observations (which are necessary e.g. if they are to be applied as discriminatory variables in econometric analysis). Still, this critique does not apply if one uses only the five categories at the highest hierarchical level, where the broad category of human capital-intensive goods substitutes for all the distinctions of high- or medium-tech production. The major shortcoming of Schulmeister’s taxonomy therefore lies in the neglect of statistical cluster techniques, which would have been a much more convenient and transparent tool for grouping observations according to multidimensional data.

The application of statistical cluster techniques is precisely what distinguishes the taxonomy presented in *Neven* (1994). Similar to the other studies, the purpose of his analysis is to “evaluate the factor content of trade in order to infer relative factor endowments and associated comparative advantage“ (ibid., p. 39). Statistical cluster techniques are applied “to define homogenous groups of industries in terms of factor intensities“ (Neven, 1996, p. 8). Four variables are used: labour intensity (the shares of wages in value added), capital intensity (the share of investment in value added), wage levels (average total compensation per worker) and human capital (the share of blue-collar workers in the total number of employees). Based on data for the years 1985 to 1990, he classifies 140 industries into five groups. Having repeated the same calculation for all EU member countries, he is able to claim that the “industry clusters are remarkably stable across countries“ (ibid., p. 8f). Subsequently, the

German classification is taken as the general reference. Despite the additional claim that the groups would be easy to interpret, Neven refrains from attaching any labels to them. They are simply numbered as industry clusters ‘1‘ to ‘5‘. The five clusters are characterised as (‘1‘) labour intensive, high-wage, and high share of white-collar labour; (‘2‘) labour intensive, high-wage, but low capital investment; (‘3‘) labour intensive, low-wage, low investment, and high share of unskilled labour; (‘4‘) capital intensive, low-wage, and a high share of blue-collar labour; and finally (‘5‘) capital intensive with high shares of white-collar labour (mostly comprised of food processing industries).

Among the many industry classifications screened for this survey, the Neven taxonomy was the first that explicitly relied on statistical cluster analysis. It therefore marks an important methodological advance. The most critical aspect of this taxonomy is that it does not rise to its full potential. As it appears, the taxonomy is purely instrumental to the analytic purposes pursued in the initial paper (and a later publication by Neven and Wyplosz, 1996), but does not induce any additional attention to or discussion of its methodological aspects. The fact that industry types are not labelled also indicates a lack of specific interest in the classification as such, as well as a common difficulty in providing an appealing economic interpretation of the clusters.

5. Intangible sources of industrial performance

Drawing on John Sutton’s (1991) theoretical concept of endogenous sunk costs, *Davies, Lyons et al.* (1996) launch a major empirical investigation into the influence of different competitive mechanisms on concentration, integration, diversification and multinationality. Emphasising the importance of product differentiation not only in industrial organisation but also in international trade theory and business economics, they develop a new taxonomy which groups industries according to the intensity of intangible and largely sunk investment into R&D and advertising. Data for R&D were taken from UK and Italian industry statistics, while advertising outlays were collected from a UK advertising agency and then aggregated to NACE 3-digit industries. Applying a conventional cut-off procedure, they group industries into four separate categories. 'Type 1' industries are characterised by the production of homogenous goods with little expenditures on R&D or advertising. 'Type 2' industries produce differentiated goods and can be further divided into three subclasses: 'Type 2A' industries with high advertising outlays; 'Type 2R' industries with high R&D expenditures; and finally 'Type 2AR' industries distinguished by high investment into both advertising and R&D.

The approach taken by Davies, Lyons et al. constitutes a significant departure from the traditional focus on tangible factors of production. Focusing on R&D and advertising as a particular intangible investment into firm specific competitive advantages, they helped to make the latest theoretical developments in industrial organisation more operational. They systematically applied a broader view of the production function in terms of all the factors deliberately used to generate revenues (instead of mere tangible output). Similar to the taxonomy developed by Neven (1996), the innovative content of the classification is somewhat detracted by the rather short and sketchy documentation (less than a page in the appendix). Understandably again, their main research interest is focused on the analysis (and on the very special data which they produced) and not on the industry classification as such. This might also explain why (in contrast to Neven) they only apply the simple cut-off approach in the identification process.

The fact that the taxonomy of Davies, Lyons et al. relates to an old NACE nomenclature initiated a series of new classifications, created on behalf of the European Commission and documented in *Peneder* (2001, 2002). The primary objective was to facilitate inquiries into industrial performance with respect to the intangible sources of competitive advantage. In later stages, they also proved to be useful tools for further analytic purposes.[11] All three taxonomies were developed by means of statistical cluster analysis. Taxonomy I focuses on the distinction between tangible and largely location-bound versus intangible and firm specific factors of production. The clustering process makes use of US data for wages and salaries, investments in physical capital, advertising outlays and R&D expenditures. These are assumed to span four independent dimensions of inputs for revenue generation. Ratios to total value added are calculated for wages and physical capital. Expenditures on advertising and R&D are represented by their ratio to total sales. The latter are directly derived from firm-level data. Taxonomy II is directed at the dimension of human resources and is based on occupational data, which distinguish the two types of white-collar and blue-collar workers, and then for each the shares of respectively high and low-skilled labour. These data stem from the OECD and cover employment shares for a sample of developed economies. Finally, Taxonomy III segregates industries according to differences in kind and intensity of external service inputs. It was created using US input-output tables, available at the disaggregate level of 500 industries, and reveals typical combinations of service inputs purchased via external market transactions.

One important drawback of the above taxonomies is that they are exclusively directed at the manufacturing sector. Implementing an analogous approach, *Mayerhofer* and *Palme* (2001) present an impressive series of classifications for the services sector based upon statistical cluster techniques and including, among others, the dimension of software intensity and the presence of agglomerative forces. Their major shortcoming within the context of this survey is that they are exclusively based on Austrian data and are therefore too specific for international comparative analysis.[12] *O'Mahony and Vecchi* (2002) stick even closer to the Peneder-type of taxonomies. In their attempt to detect industry specific affects on company performance, they extend the categories of Taxonomies I and II, so that they also cover non-manufacturing industries. Furthermore, they introduce an additional industry classification, discriminating according to the intensity with which ICT equipment is used. Based on various data sources from the US Bureau of Labour Statistics, company accounts and the Euromonitor marketing year book, their additional identification of 24 two-digit non-manufacturing sectors is, in the case of Taxonomy I, based on an ad hoc combination of several cut-off procedures. For the human resources dimension, however, statistical clustering is applied in the division of 139 industries into three groups of high, intermediate and low skilled labour, based on data from the UK labour force survey. The final taxonomy for ICT investment is based on US BEA data covering 44 industries for the years 1988 to 1997.

At this point, however, an opposing view should also be mentioned. In their discussion of the principal usefulness of extending Peneder's taxonomic approach to the services sector, *Miles* and *Tomlinson* (2000) take a sceptical stance. In short, they conclude that "approaches to clustering developed for manufacturing are unlikely to be reproduced for services" (ibid., p.181). Apart from the frequent lack of data, they argue that the large heterogeneity within service activities is most problematic. On a conceptual basis, they favour a combination of the traditional functional separation by Singelmann (1979) into (i) producer services, (ii) distributive services, (iii) personal services, and (iv) social services, together with a different typology of services in terms of their core processes.[13] The latter distinguishes services as (i) 'effecting physical transformations'; (ii) 'those directly concerned with changing the (biological or psychological) state of human beings'; and (iii) 'those whose main activities involve information' (ibid., p. 158f). They nevertheless embrace quantitative approaches towards classifying industries and emphasise that several "useful measures for distinguishing between services do appear to emerge, ... especially, those relating to the knowledge-intensity of their activities, as indicated by such variables as skill composition and investment in development and technological knowledge, and the use of IT and IT services. Other parameters of intangible investment, such as advertising spending, do not appear to form the basis of distinct clusters" (ibid.).

Miles and Tomlinson thus suggest some important directions for future research. A first major initiative in this direction was undertaken by *van Ark* (2001), who complemented an OECD definition of 'ICT-producing sectors' with a separate class of 'ICT-using industries'. The latter was identified by a cut-off procedure applied to the share of ICT investment in output and the industry share in ICT capital stock. The data are for the Netherlands and the USA. The classification refers to ISICrev.3 2-digits with some corrections for a few 3-digit industries. ICT-producing industries were excluded from this category. The residual of all other industries was labelled the 'Non-ICT Sector'.

6. Quality competition

Aiginger (2001) puts forward an innovative and entirely distinct classification, which tries to capture characteristic differences between industries with respect to their typical reliance on quality versus price competition. Exploring the explanatory power of unit values in foreign trade statistics, which are calculated as the nominal value of trade flows divided by their physical quantities (usually in tons or kilogram), he calculated the net balance of exports minus imports for both unit values and the physical quantities traded. The data are from the bilateral trade flows of all the EU member countries vis-à-vis 30 trade partners (EU, USA, Japan, 8 NICs and 6 accession countries) at a very detailed level, which are aggregated to NACE 3-digit industries. He then applies the following logic for identification: "Industries in which higher prices (more exactly: higher unit values in exports relative to imports) are associated with lower quantities (more exactly: lower exported quantities relative to imported quantities) are revealed to be price elastic. Industries in which the signs of (net) prices and (net) quantities are the same are revealed to be quality elastic" (ibid., p. 20). The overall share of identical signs is then taken as an indication of the importance of quality as opposed to price competition. Empirically, the share of positive signs lies between 25.0 % for the most price sensitive and 53.5 % for the most quality sensitive industries. Aiginger's next step was to apply a cut-off procedure, which separates all the NACE 3-digit manufacturing industries in three equally-sized groups of (i) high-, (ii) medium-, and (iii) low revealed quality elasticity (RQE).

A distinctive feature of Aiginger's approach lies in its application of a simple data concept to the extremely difficult question of how to measure product quality. The value of the classification therefore stands and falls with the accuracy of how unit values can serve this purpose. Abounding conceptual and measurement problems assure us that we can not expect any perfect matching. Aiginger is very explicit about the shortcomings and stresses that some industries, such as textiles, clothing, footwear or jewellery, have intrinsically higher unit values without any obvious quality mark-ups relative to other goods (ibid., p. 13). In addition, we must keep in mind that unit values primarily reflect the number of processing stages already in operation and therefore tend to rise as one moves along the value added chain. Product quality is then equivalent to the amount of non-material inputs added to the various stages of production and is further augmented by the more initial material inputs which are consumed during the process. Although products thus tend to have higher unit values at higher stages, the same need not apply to the particular value that is added at each stage of processing. But keeping these caveats in mind, unit values are still the most comprehensive measure of product quality at hand. Invoking various activities that increase the quality of manufactured products (e.g. increasing durability and reliability; using superior inputs and skills; making the product more specific to demand; adding new functions, better design, advertising, etc.; ibid., p. 11), Aiginger argues that these would all lead to an increase in the value of products relative to their initial physical quantity. If we then add the practical advantage of large and internationally comparable data at a very disaggregate level, unit values are indeed a promising output-related measure for creating industry classifications.

7. Summary and conclusions

In this interpretative survey we have screened numerous industry classifications, deemed to be of relevance for applied analysis in industrial economics, technological development, international trade, and competitive performance. Our intention was to provide a brief overview of relevant approaches to the practitioner in the field of applied empirical studies. A surprisingly large variety of available taxonomies, directed at different analytic questions and relying on various data, nomenclatures, and methods of identification surfaced. In principle, this diversity is welcome, as it offers researchers a wide choice. In practice, however, the range of taxonomies is greatly limited by either a lack of information and precise documentation or their adherence to outdated official nomenclatures. Given the fundamental importance of industry classifications in investigating the impact of the characteristic market environment on economic activity, we conclude that the much neglected process of creating classifications deserves more attention within the discipline of economics and should be pursued on a more systematic and professional basis. Ideally, the overall target is to make available a variety of classifications, which are commonly based on state of the art methodology, designated by the currently most important international nomenclatures, and available at different levels of aggregation.

Despite the diversity of the approaches, several problems and caveats are held in common. Above all, when implementing a taxonomic approach, one must stay alert of the hidden heterogeneity, which regularly prevails within the categories of a certain type. One generally would not pursue a taxonomic approach for purely analytic purposes, if more disaggregated data (or even micro data) were available in a comparable format. One would generally prefer to work with the latter and benefit from the richer set of available information. Industry classifications are therefore of greatest value, if they reflect systematic relationships for data which otherwise are difficult to compare. Because an exact correspondence is not necessary for applying the classification as a discriminatory variable, the only requirement for assuming stability between countries (or over time) is the consistency of the assignments with respect to broadly defined industry types.

The actual validity of stability assumptions depends of course on the variable used for identification. Some of them might be very characteristic and stable at the industry level, others will be more idiosyncratic and might lack a systematic relationship. But using the simple and therefore popular cut-off approach, any arbitrary chosen variable can be applied in the classification of observations. No alarm-bell rings, if the variable fails to reflect any systematic structural relationship. In order to raise the state of the art within our discipline, two almost obvious requirements should therefore become imperative. First, good classifications need to be explicitly guided by economic theory, telling us where to expect systematic relationships at the level of markets and technologies. Secondly, statistical cluster techniques should be regularly applied by the researcher as he seeks to explore and discover the structure within the data.

Apart from these general requirements, the desired mitigation of problems from hidden heterogeneity also sets the agenda for the most promising routes for the future development of industry classifications.

(i) First of all, the fundamental heterogeneity within industries implies that the accurateness of a taxonomy crucially depends on the level of disaggregation. In many instances, more disaggregated data would be available, but are not published for reasons of confidentiality. Closer co-operation between statistical offices and the research community is therefore desirable.

(ii) Secondly, many classifications are based only on data for one country. Due to the potential effects of heterogeneity between different economies, a broadening of the geographic coverage together with tests on the significance of between country variation is needed. In general, data from large economies that are less exposed to distortions by local idiosyncrasies are preferable to data from small countries.

(iii) Thirdly, the integration of services into the more elaborate taxonomies for the manufacturing sector remains a critical task. The lack of equally disaggregated data and the large degree of heterogeneity within the services sector, together with the wide range of differences in technology and the market environments with respect to manufacturing and services, suggest that in many instances separate classifications would be more appropriate. An integrated approach might nevertheless remain desirable for higher levels of sectoral aggregations.

Acknowledgements

It has been my great fortune to partake in numerous conversations about industry classifications and their application during the past years. My especial thanks go to Karl Aiginger, Birgit Andersen, Steve Davies, Giovanni Dosi, Manfred Fischer, Lionel Fontagné, Paul Geroski, Heinz Hollenstein, Alexis Jacquemin, Serguei Kaniovski, Bruce Lyons, Mary O'Mahony, Peter Mayerhofer, Stan Metcalfe, Ian Miles, Alessandro Sembenelli, Michael Pfaffermayr, Stephan Schulmeister, Gunter Tichy, Michela Vecchi, and Reinhilde Veugelaers.

Notes

[1] Consequently, we will not consider official nomenclatures applied in the statistical offices, which are necessarily geared according to more practical considerations such as the ease of systematic data collection and comprehensive coverage. For recent debates related to the introduction of the new North American Industry Classification System (NAICS) see Murphy (1998) or Kort (2001).

[2] Economic Geography is one sub-discipline for which this general assessment does not apply. For example, Fischer (1979) offers a detailed discussion of the various methods to create regional classifications.

[3] For the sake of simplicity, we will not distinguish between the categorisation of industrial activities and the categorisation of product groups, but will refer to both as industry classifications. However, we present information on the respective nomenclatures for which the classifications have been built, thus indicating, for example, whether it is available for the product groups in trade statistics (such as SITC or HS/CN) or the activity based schemes in common production statistics (such as ISIC, USSIC or NACE).

[4] In the OECD classification, knowledge-based services comprise communications (ISIC 72), finance, insurance, real estate and business services (ISIC 8), as well as community, social and personal services (ISIC 9).

[5] "[T]he distinction between the two Schumpeterian regimes involves a reversal of the relative roles of innovation by entrants and established firms. An entrepreneurial regime is one that is favourable to innovative entry and unfavourable to innovative activity by established firms; a routinised regime is one in which conditions are the other way round" (Winter, 1984, p. 297). Malerba and Orsenigo (1996) apply this distinction in an empirical setting under the headings of Schumpeter Mark I and Mark II industries.

[6] "They appropriate less on the basis of a technological advantage, than of professional skills, aesthetic design, trademarks and advertising. Technological trajectories are therefore defined in terms of cutting costs" (Pavitt, 1984, p. 356)

[7] Her technological regimes are characterised along the following dimensions: technological opportunity; technological entry barriers; persistence of innovation; inter-firm diversity; differentiation of the knowledge bases; relevance of external knowledge sources; links with academic systems; and finally the nature of innovation (i.e. process versus product innovation).

[8] For a conceptual typology of services in reply to Pavitt (1984) see Soete and Miozzo (1989).

[9] He then developed a corresponding classification, which is based on 52 ISIC manufacturing industries for the USA and, among others, uses data from 1972 input-output tables. A related hybrid taxonomy, largely disappointing in its attempted combination of various older classifications, was presented by *Robert Ballance* (1992).

[10] The cut-off procedure produces the value ‘1‘ if the according intensity for an industry is more then 10 % above the German average and is otherwise ‘0‘. Based on subjective expert opinions, the variable for research intensity was then further divided into the two classes of high- and medium-tech. Finally, dependence on raw materials was separated into agricultural and mineral sources.

[11] Probably the best one can say about an industry classification is to report its successful applications. In the case of the above taxonomies, Kaniovski and Peneder (2002) demonstrate by means of binomial tests and the decomposition of variances that these are indeed a useful and reliable tool in the identification of a statistically significant structural dimension of competitive strategy. In a complementary study, Peneder (2002) investigates the meso-macro link and finds that specific categories of Taxonomies I and II have a significant influence upon aggregate income levels and growth in a dynamic panel econometric framework.

[12] Aiming at the analysis of Austrian concerns, this is considered to be an advantage within the context of the study by Mayerhofer.

[13] This second typology was initially set out in Miles (1993). Finding an appropriate theoretical characterisation of services still raises enormous conceptual difficulties, as recently demonstrated in Gadrey (2000) and Hill (1999).

References

Aiginger, K., Europe's Position in Quality Competition. Background Report for "The European Competitiveness Report 2000," Enterprise Papers No. 4, European Communities, Luxembourg, 2001.

Archibugi, D., "Pavitt's Taxonomy sixteen years on: a review article", Economics of Innovation and New Technology, vol. 10, pp. 415 – 425, 2001.

Arvanitis, S., Hollenstein, H., Innovative Activity and Firm Characteristics – A Cluster Analysis with Firm-level Data of Swiss Manufacturing, paper presented at the 25th Annual EARIE Conference, Copenhagen, August 27-30, 1998.

Bailey, K.D., Typologies and Taxonomies. An Introduction to Classification Techniques, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-102, Thousand Oaks, CA: Sage, 1994.

Ballance, R., European Economic Restructuring: Retrospect and Prospect, in: Cool, K., Neven, D.J., and Walter, I., European Industrial Restructuring in the 1990s, Houndmills, UK: Macmillan, pp. 25-56, 1992.

Capelot, E.B., Lambertz, J.E., Hi-tech products, paper presented at an OECD Seminar on High-Technology Industry and Products Indicators, Paris, 25th to 26th November, 1993.

Cesaratto, S., Mangano, S., "Technological Profiles and Economic Performance in the Italian Manufacturing Sector", Economics of Innovation and New Technology, vol. 2, pp. 237-256, 1993.

Davies, S. and Lyons, B. (eds.), Industrial Organization in the European Union. Structure, Strategy, and the Competitive Mechanism, Clarendon Press: Oxford, 1996.

Davies, S., Rondi, L., and Sembenelli, A., "European Integration and the Changing Structure of EU Manufacturing, 1987-1993", Industrial and Corporate Change, vol. 10 (1), pp. 37-75, 2001.

Dosi, G., "Technological Paradigms and Technological Trajectories", Research Policy, 11, pp. 147-162, 1982.

Dosi, G., "Sources, Procedures, and Microeconomic Effects of Innovation", Journal of Economic Literature, 26 (3), pp. 1120-1171, 1988.

Dosi, G., Malerba, F. (eds.), Organization and Strategy in the Evolution of the Enterprise, MacMillan: London, 1996.

Evangelista, R., "Sectoral Patterns of Technological Change in Services," Economics of Innovation and New Technology 9, pp. 183-221, 2000.

Fischer, M. M., "Regional Taxonomy. A Comparison of Some Hierarchic and Non-Hierarchic Strategies," Regional Science and Urban Economics, 10, pp. 503 – 537, 1980.

Gadrey, J., “The Characterization of Goods and Services: An Alternative Approach,” Review of Income and Wealth, 46 (3), pp. 369-387, 2000.

Hill, T.P., “Tangibles, Intangibles and Services: A New Taxonomy for the Classification of Output”, Canadian Journal of Economics, 32 (2), pp. 426-46, 1999.

Hollenstein, H., Patterns and Determinants of International Activities: Are SMEs Different?, WIFO Working Papers 156, Vienna, 2001.

Kaniovski, S., Peneder, M., “On the Structural Dimension of Competitive Strategy”, forthcoming in: Industrial and Corporate Change, vol. 11 (3), 2002.

Kort, J. R., “The North American Industry Classification System in BEA’s Economic Accounts,” Survey of Current Business, 81 (5), pp. 7 – 13, 2001.

Lawrence, R.Z., Can America Compete?, The Brookings Institution, Washington D.C., 1984.

Lazarsfeld, P.F., “Some Remarks on Typological Procedures in Social Research”, Zeitschrift für Sozialforschung 6, pp. 110-139, 1937.

Malerba, F., Orsenigo, L., “Technological Regimes and Firm Behavior”, Industrial and Corporate Change, 2 (1), pp. 45-71, 1993.

Malerba, F, Orsenigo, L., “Schumpeterian Patterns of Innovation are Technology Specific”, Research Policy, 25, pp. 451-478, 1996A.

Marsili, O., The Anatomy and Evolution of Industries. Technological Change and Industrial Dynamics, Edward Elgar: Cheltenham, 2001.

Mayerhofer, P., Palme, G., Sachgüterproduktion und Dienstleistungen: Sektorale Wettbewerbsfähigkeit und regionale Integrationsfolgen (Preparity, Teilprojekt 6/1); WIFO, Vienna, 2001.

Mc Guckin, R.H., Abbott, T.A., Herrick, P., and Norfolk, L., Measuring Advanced Technology Products Trade: A New Approach, paper presented at an OECD Seminar on High-Technology Industry and Products Indicators, Paris, 25th to 26th November, 1993.

Menger, C., Grundsätze der Volkswirtschaftslehre, Vienna: Hölder-Pichler-Tempsky, 2nd edn, 1871/1923.

Miles, I., “Services in the New Industrial Economy”, Futures, 25 (6), pp. 653-672, 1993.

Miles, I., Tomlinson, M., Intangible Assets and Service Sectors: The Challenges of Service Industries, in: Buigues, P., Jacquemin, A. and Marchipont, JF, Competitiveness and the Value of Intangible Assets, Edward Elgar, Cheltenham, UK, pp. 154-186, 2000.

Murphy, J.B., “Introducing the North American Industry Classification System”, Monthly Labour Review, 121 (7), pp. 43-47, 1998.

Nelson, R.R. and Winter S.G., An Evolutionary Theory of Economic Change, Belknap Press: Cambridge, Massachusetts, 1982.

Neven, D., “Trade Liberalisation with Eastern Nations, How Sensitive?“, in Faini, R. and Portes R., European Union Trade with Eastern Europe, Adjustment and Opportunities, CEPR, London, 1994.

Neven, D., Wyplosz, C., “Relative Prices, Trade and Restructuring in European Industry“, CEPR Discussion Paper, No. 1451, 1996 (also published in: Dewatripont, M, Sapir, A., Sekkat, K. (eds.), Trade and Jobs in Europe: Much Ado about Nothing?, Oxford: Oxford University Press, pp. 33-59, 1999).

OECD, Science, Technology and Industry Scoreboard 1999. Benchmarking Knowledge-Based Economies, Paris: OECD, 1999.

O’Mahony, M., Vecchi, M., Knowledge based capital and company performance, mimeo, 2002.

Peneder, M., Industrial Structure and Aggregate Growth, mimeo, 2002B.

Peneder, M., Intangible Investment and Human Resources, The Journal of Evolutionary Economics, 12 (1-2), pp. 107-134, 2002A.

Peneder, M., Entrepreneurial Competition and Industrial Location, Edward Elgar, Cheltenham, UK, 2001.

Peneder, M., Intangible Assets and the Competitiveness of European Industries, in: Buigues, P., Jacquemin, A., Marchipont, F. (eds.), Competitiveness and the Value of Intangible Assets, Edward Elgar, Cheltenham, 2000.

Schulmeister, S., “Das technologische Profil des österreichischen Außenhandels”, WIFO-Monatsberichte, 63 (12), pp. 663-675, 1990.

Soete, L., Miozzo, M., Trade and Development in Services: A Technological Perspective, Working Paper No. 89-031, MERIT, Maastricht, 1989.

Streissler, E.W., “Structural Economic Thought. On the Significance of the Austrian School Today”, Zeitschrift für Nationalökonomie, 29 (3-4), pp. 237-266, 1969.

Van Ark, B., The Renewal of the Old Economy: An International Comparative Perspective, mimeo, 2001.

Winter, S.G., “Schumpeterian Competition in Alternative Technological Regimes”, Journal of Economic Behavior and Organization, 5 (3-4), pp. 287-320, 1984.

Table 1: Summary of the main characteristics of the taxonomies screened (in chronological order of publication)

Industry classifications	Initial aim	Data	Methodology	Final categories
Pavitt (1984)	To "describe and explain the sectoral patterns of technical change" (p. 343)	2000 significant innovations, and innovating firms", UK; 1945 to 1979; 3- and 4-digits (half of manufacturing output);	Inductive identification of reduced set of quantitative and qualitative information;	Sectors with many (i) science based firms; (ii) productionintensive: either scaleintensive or specialised suppliers; (iii) supplier dominated;
Schulmeister (1990)	Mapping of technology profile from patterns of international trade specialisation	Factor intensities from Legler (1984); Germany, ca. 1980; SITC rev2 3-digits for manufacturing	Inductive identification from multidimensional binary data (cut-off)	Multilevel hierarchy, combi-ning: high-skills; high- and medium-tech; capital-, labour-, and resource-intensive goods
McGuckin et al. (1992)	To measure high-tech products in US and international trade	US SIC, high disaggregation	Identification by subjec-tive expert opinion	Single separation of 'advanced technology products'
Capelot and Lambertz (1993)	To measure high-tech products in EU and international trade	HS 6-digits	Identification by subjec-tive expert opinion	List of 'high-tech' products
Neven (1994)	To "evaluate the factor content of trade" (p. 39)	Labour, capital intensity, wage levels and share of blue-collar workers; 140 NACE 3- and 4-digits; EU members; 1985-90	Quantitative identific-ation by statistical cluster analysis	Five clusters characterised by various combinations of the above variables; no simple labels offered
Davies and Lyons (1996)	To investigate the influence of compe-titive mechanisms on concentration, integrat-ion, diversification and multinationality;	UK advertising agency data for 1987 aggregated to NACE 3-digits; R&D ratio for UK and Italy; NACE 2 or 3-.digits	Quantitative identificat-ion by cut-off procedure	(1) homogenous goods; (2) differentiated goods: (2A) advertising-; (2R) R&D-; (2AR) R&D and advertising-intensive industries

Table 1: Summary of the main characteristics of the taxonomies screened (continued)

Industry classifications	Initial aim	Data	Methodology	Final categories
Hatzichrinoglou (1997) I. High-tech sectors	To establish an authoritative list of high-tech industries for international studies	Direct + indirect R&D intensity in OECD countries; 1980 to 1990; / 22 ISIC rev2 2-digits (manufacturing only)	Quantitative identific-ation by the cut-off procedure	One-dimensional split into 4 groups: high; medium-high; medium-low; and low technology sectors
Hatzichrinoglou (1997) II. High-tech products	To establish an authoritative list of high-tech products for international studies	R&D intensities for selected OECD members in co-operation with ISI (Germany); 1985 to 1990; / 126 SITC rev3 5-digits	Quantitative identifi-cation (cut-off) modified by subjective expert opinions	Single separation of high technology products
Evangelista (2000)	To provide "a comprehensive picture of the characteristics of innovation in services" (p. 183)	11 innovation related variables aggregated by factor analysis; Italian innovation survey, 1993 to 1995; firm-level aggregated to 22 service sectors	Quantitative identifi-cation by factor analysis and cluster techniques	Services separated into (i) tech-nology users; (ii) science & technology-based; (iii) interac-tive- & IT-based; (iv) technical consultancy
Peneder (2000) (External services)	To investigate intangible sources of competitive advantage; meso-micro and meso-macro link;	Principle components of various service inputs; US input-output table (ca. 500 industries); US SIC, 1992	Quantitative identifi-cation by principle com-ponent and statistical cluster analysis	Industries with high inputs from (i) knowledge-based services; (ii) retail & adverti-sing; (iii) transport; (iv) other
Marsili (2001)	To investigate technological change and industrial dynamics	Various sources of SPRU databases on innovation and innovative firms	Identification based on quantitative and quali-tative information	Technological regimes: (i) sci-ence based; (ii) fundamental processes; (iii) complex sys-tems; (iv) product engineering; (iv) continuous processes

Table 1: Summary of the main characteristics of the taxonomies screened (continued)

Industry classifications	Initial aim	Data	Methodology	Final categories
van Ark (2001)	To investigate the contribution of ICT to output and productivity growth	ICT investment and capital stock; ISICrev3 2-digits; Netherlands and the USA, approx. 1990 to 1995/98;	Quantitative identific-ation by cut-off procedure	Separating (i) 'ICT-using industries' complementary to (ii) 'ICT-producing industries' and (iii) 'others'.
Aiginger (2001)	To "investigate how Europe is positioned in quality competition" (p.2)	Unit values; bilateral trade of EU members to 30 countries, 1988; CN 6-digit	Quantitative identific-ation by cut-off procedure	Revealed quality elasticity (RQE): (i) high; (ii) medium; and (iii) low
Peneder (2001) (Intangible investment)	To investigate intangible sources of competitive advantage; meso-micro and meso-macro links	Labour and capital intensity; share of R&D and advertising in total turnover; USA; NACE 3-digits; 1994/1995	Quantitative identific-ation by statistical cluster analysis	Industries characterised as (i) technology driven; (ii) marketing driven; (ii) labour-; (iii) capital-intensive; (v) mainstream
Peneder (2001) (Human resources)	To investigate intangible sources of competitive advantage; meso-micro and meso-macro links	Employment shares of high- or low-skilled; white- or blue collar labour; selected OECD countries; 1990 to 1994; ISIC 2-digits	Quantitative identific-ation by statistical cluster analysis	(i) High-skilled industries; (ii) medium-skilled white-collar; (iii) medium-skilled blue-collar; (iv) low-skilled
O'Mahony and Vecchi (2002)	To investigate the effects of knowledge-based activities on company performance	Various data sources; disaggre-gation: between 22 and 139 industries	Quantitative identific-ation by cut-off or statistical clustering	Extending categories in Peneder (2001) to non-manu-facturing; additional separation of 'ICT-intensive sectors'