Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

Discussion Papers

884

Jens Schmidt-Ehmcke • Petra Zloczysti

Research Efficiency in Manufacturing – An Application of DEA at the Industry Level

Berlin, May 2009

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the institute.

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Available for free downloading from the DIW Berlin website.

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN.
Papers can be downloaded free of charge from the following websites:

http://www.diw.de/english/products/publications/discussion_papers/27539.html
http://ideas.repec.org/s/diw/diwwpp.html
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=1079991

Research Efficiency in Manufacturing

-

An Application of DEA at the Industry Level[1]

by

Jens Schmidt-Ehmcke (DIW-Berlin)*
Petra Zloczysti (Freie Universität Berlin)**

May 2009

* German Institute for Economic Research, Mohrenstr. 58, 10117 Berlin, jschmidtehmcke@diw.de

** John F. Kennedy Institute for North American Studies, Lansstr. 7-9, 14195 Berlin petra.zloczysti@fu-berlin.de

[1] This paper is a result from a research program on innovation economics at the DIW Berlin. We thank without implying Irwin Collier, Astrid Cullmann, Christian von Hirschhausen, Alexander Kritikos and Andreas Stephan for their support and helpful comments.

Abstract

This paper analyzes research efficiency at the industry level in manufacturing for 13 European member and four nonmember countries during 2000 and 2004. A unique dataset was compiled that matches patent applications at the European Patent Office (EPO) to industry-specific R&D inputs from EU KLEMS. We find that Germany, the United States, and Denmark have the highest efficiency scores on average in total manufacturing. The main industries that are at the technology frontier are those involved in electrical and optical equipment and machinery. Separate frontier estimations for these industries, conducted without the constraint of a constant technology frontier, provide additional support for our results.

Keywords: R&D efficiency, industry level, data envelopment analysis, manufacturing

JEL Classification: C14, L60, O31, O57

1 Introduction

A knowledge-production function is central to many endogenous economic growth models in which innovation plays a crucial role in sustaining long-term growth. Innovation becomes even more important to productivity growth when a particular national industry approaches the world technology frontier because, at that point, imitation, as opposed to true innovation, is less feasible. Empirical literature confirms the importance of research and development (R&D) expenditures to economic growth [e.g., GUELLEC & VAN POTTELSBERGHE DE LA POTTERIE, 2001]. The resources available for the generation of new knowledge are often limited and thus need to be used as efficiently as possible to sustain and promote long-term growth.

Our paper aims at identifying the country-industry combinations that define the world technology frontier in the manufacturing sector. In the literature to date, country-level studies assume a common technology frontier across all industries under observation. Obviously, however, calculating efficiency at the country level ignores differences in the structure and efficiency of different industries. This paper intends to discover which countries have the most efficient industry-specific knowledge production processes. First, we derive efficiency estimates for the entire manufacturing sector at the country level. Second, we relax the assumption of a common country-industry technology frontier and identify those county-industry combinations that are occupying the world technology frontier. Third, we focus on those industries with the highest efficiency scores — that is, the industries that define the technology frontier — and conduct separate efficiency analyses to add further solid support to our results.

Identifying the best-performing industries among countries can serve the useful purpose of providing a benchmark against which other industries' strengths and weaknesses can be measured. Being able to conduct a performance assessment of knowledge production will help decision makers allocate limited financial resources efficiently so as to achieve the most knowledge production possible. In addition, countries with less efficient industries can use our findings regarding the most efficient countries to improve their own processes.

Although a number of studies measure research efficiency at the country level, ours is the first to analyze it at the industry level. This focus on the industry level of knowledge production provides detailed insight into efficiency differences within and across countries' research activities. It allows us to conduct a fine-grained examination of various nations' domains of specialization, measured by a high share of gross output and industries occupying the world technology frontier.

Our study is based on a unique industry dataset compiled from two sources: EU KLEMS and PATSTAT. We match EPO patent applications to the EU KLEMS industry-level data by using the concordance provided in SCHMOCH & AL. [2003]. To our knowledge, this paper is the first to link these two sources, thus making a unique contribution to the study of research efficiency.

To measure research efficiency across industries, we employ the nonparametric DEA method, an approach well suited, for several reasons, for measuring R&D performance [WANG & HUANG, 2007]. It requires no specification of the functional form of the knowledge production process; neither does it need any a priori information concerning the importance of inputs and outputs. Since DEA is a deterministic approach, extreme observations can have a strong influence on the calculated efficiencies. We circumvent this problem by using the super-efficiency approach of BANKER & CHANG [2006] to detect and then remove extreme observations from the sample, thus achieving a consistent and robust technology frontier. Furthermore, industries of various economic sizes are compared in our model. It is both statistically and economically important to determine whether the underlying technology exhibits increasing, constant, or decreasing returns to scale. Therefore, we test the hypotheses of constant returns to scale using the bootstrap procedure proposed by SIMAR & WILSON [2002].

Our paper is organized as follows. Section 2 introduces the analytical framework and briefly summarizes the literature in this field. In Section 3, the methodology of data envelopment analysis (DEA) studies is introduced. Section 4 describes the model specification and data. The empirical results for total manufacturing and by industry are presented in Section 5. Section 6 summarized the findings and concludes.

2 Analytical Framework

We focus on the economic process leading to reduction in the cost of producing existing products (process innovations) or in the development of new products (product innovations). In particular, we analyze whether there are substantial differences in knowledge creation between countries and industries.

Our model follows the "knowledge production function" framework first articulated by GRILICHES [1979] and implemented by PAKES & GRILICHES [1984], JAFFE [1986], HALL & ZIEDONIS [2001], among others. Innovative output is the product of knowledge-generating inputs, similar to the production of physical goods. Some observable measures of inputs, such as R&D expenditure and high-skilled labor and researchers, are invested in a knowledge production function. These "inputs" are directed toward producing economically valuable knowledge. The production process is viewed as a continuum leading from R&D and human capital (the inputs) to some observable measure of innovative activity:

$$I_{ci} = f(R\&D_{ci}, HS_{ci}, MS_{ci})$$

where I is innovative output, *R&D* denotes the R&D capital stock as a proxy for accumulated knowledge, and *HS* and *MS* are, respectively, the number of high-skilled and medium-skilled workers employed. The unit of observation is the country (c) industry (i) level.

Innovative output as the result of knowledge production is difficult to measure. We use patent applications as a measure of successful knowledge production, although doing so has its drawbacks. First, patent applications are often criticized as measuring just one component of the innovative output since inventors may choose other protection strategies, such as trade secrets. Thus, the use of patents underestimates real innovative activity. Second, research [E.G., SCHERER, 1965; PAKES & SCHANKERMAN, 1984; PAKES, 1986; GRILICHES, 1990] shows that the value of patents is skewed to the right, with only a few patents being highly valuable. Despite this criticism, however, patents are probably the best indicator of research output and are widely used as such in the literature [E.G., HAUSMAN ET AL., 1984; KORTUM, 1997; TEITEL, 1994]. First, they are by definition related to inventiveness and based on an objective and fairly time-insensitive standard. Second, data on patent applications are widely available and provide additional information about the origin of the inventor and a detailed technological classification of the underlying invention.

Based on a knowledge production function framework, the literature confirms the importance of research personnel and R&D capital to the knowledge creation process; however, far less attention has been paid to the importance of the efficient use of scare resources in this process.

ROUSSEAU & ROUSSEAU [1997, 1998] were the first to use a DEA approach to assess the relative efficiency of the R&D process. Using a sample of 18 developed countries, they applied an input-oriented, constant return to scale model with two outputs—the number of scientific publications and the number of granted patents at the European Patent Office (EPO) — and used GDP, along with population and R&D investment, as input factors. Based on their data and specification, they found Switzerland to be the most efficient country in Europe in 1993, followed closely by the Netherlands. Using the same framework, ROUSSEAU & ROUSSEAU [1998] extended their work on R&D efficiency by including the non-European countries, specifically the United States, Canada, Australia, and Japan. With the caveat that the findings could contain some bias due to using EPO patent applications for the non-European countries, the authors reaffirmed their previous conclusion that Switzerland, again followed by the Netherlands, are the countries with the highest research efficiency.

LEE & PARK [2005] measure R&D efficiency in 27 countries with a special emphasis on Asia. They expand ROUSSEAU & ROUSSEAU'S [1997, 1998] basic framework by using the technology balance of receipts as an additional output of the innovation process. In their basic model, Austria, Finland, Germany, Hungary, and Great Britain are found to occupy the technology frontier.

WANG & HUANG [2007] propose a three-stage approach to evaluating the relative technical efficiency of R&D across 30 OECD member and nonmember countries that controls for cross-country variation in external factors such as the enrollment rate in tertiary education, PC density, and English proficiency. In the first stage, they apply an input-oriented DEA analysis where patents and publications serve as outputs and R&D expenditure and researchers as inputs. Their findings indicate that about half the countries in their sample are efficient in R&D activity. In a second stage, they take the input slacks generated in the first stage as the dependent variable for a Tobit regression in order to purge external effects caused by environmental factors outside the efficiency evaluation. Using the results from the second stage, an additional DEA is conducted, the results of which indicate a decrease in the number of efficient countries due to the external factors.

A recent study by SHARMA & THOMAS [2008] measures the efficiency of the R&D process across 18 countries using a DEA approach that applies constant as well as variable returns to scale production technology. Their approach deviates from previous work in two ways. First, they consider a time lag between R&D expenditure and patents granted and, second, they include developing countries in their analysis. Their main findings indicate that when using the constant returns to scale approach, Japan, the Republic of Korea, and China occupy the efficiency frontier, whereas within the variable returns to scale framework, Japan, the Republic of Korea, China, India, Slovenia, and Hungary are found to be efficient.

3 Methods

Data envelopment analysis (DEA) is a nonparametric approach to measuring the efficiency of a DMU that neither requires any assumptions about the functional form of a production function nor any a priori information on the importance of inputs and outputs. Central to DEA is the production frontier, which is defined as the geometrical locus of optimal production plans [SIMAR & WILSON, 1998, 2007]. Using linear programming techniques, a piecewise linear surface, or frontier, that envelopes the data is constructed as a reference point. The individual efficiencies of each DMU relative to the production frontier are then calculated by means of distance functions. DMUs located on the frontier are considered 100% efficient, whereas DMUs with efficiency scores below 100% are inefficient. The distance to the frontier is thus a measure of inefficiency. There are basically two types of DEA model: those that maximize outputs, leaving the input vector fixed (output-oriented), and those that minimize inputs, keeping the output vector constant (input-oriented).

We use the output-oriented approach with constant returns to scale technology. The efficiency score of the i th industry in a sample of N industries in the constant returns to scale (VRS) model is determined by the following optimization [COELLI & AL., 2005]:

$$\begin{aligned}
&\max_{\phi,\lambda} \phi, \\
&s.t \\
&-\phi y_i + Y\lambda \geq 0, \\
&x_i - X\lambda \geq 0, \\
&I1'\lambda = 1, \\
&\lambda \geq 0
\end{aligned}$$

where λ is a N×1 vector of constants and X, Y represent input and output vectors. ϕ measures the radial distance between the observation x_i, y_i and the efficiency frontier. The efficiency score is the point on the frontier characterized by the level of inputs necessary to be efficient [SIMAR & WILSON, 1998]. A value of $1/\phi = 1$ indicates that an industry is fully efficient and thus is located on the efficiency frontier.

Different assumptions can be made regarding the underlying technology that defines the frontier. Here, we distinguish between two types of technology: constant returns to scale (CRS) [CHARNES & AL., 1978] and variable returns to scale (VRS), which assumes that scale inefficiencies are present [BANKER & AL. 1984]. The only difference between the CRS and the VRS models is the presence of an additional convexity condition $\sum\lambda=1$.

Within this framework, industries of different sizes concerning the input requirements are compared. It is both statistically and economically important to determine whether the underlying technology exhibits increasing, constant, or decreasing returns to scale. If we assume, a priori, CRS technology without investigating the possibility that it is nonconstant, we run the risk that our efficiency estimates will be inconsistent. On the other hand, if we assume variable returns to scale when, in reality, the technology exhibits global constant returns to scale, there may be a loss of statistical efficiency [SIMAR & WILSON, 2002]. To test hypotheses regarding returns to scale we employ a bootstrap procedure. We test the null hypothesis (H_0) of a global CRS production frontier against the alternative hypothesis (H_1) that the production frontier exhibits VRS. Our test statistic is the estimated ratio between the CRS and the VRS efficiency; formally

$$\hat{\omega} = \frac{\hat{\theta}_n^{CRS}(x, y)}{\hat{\theta}_n^{VRS}(x, y)}.$$

This statistic provides an estimate of the distance between both frontiers. The appropriate p-values are calculated by means of bootstrapping.[2]

Our DEA estimator is a deterministic frontier model, which implies that all observations are assumed to be technically attainable. The main drawback of deterministic frontier models is that they are highly sensitive to outliers and extreme values in the data [SIMAR & WILSON,

2 For a detailed description of the test procedure, see Simar & Wilson [1998, 2002].

2000, 2007]. Outliers are extreme observations often caused by errors in measuring either the inputs or outputs. It is therefore important to assess ex ante whether the data contain outliers that are driving the location of the efficiency boundary and inappropriately influencing the performance estimations of the other DMUs in the sample. In this paper we use the super-efficiency method proposed by ANDERSEN & PETERSEN [1993] and BANKER & CHANG [2006] to identify and remove extreme values ex ante. The concept of super-efficiency is based on the idea of re-estimating the production frontier with different sets of observations from the sample. At every step some of the efficient DMUs are excluded from the reference set so that it is possible to obtain efficiency scores that exceed 1. If an efficient observation is an outlier, it is more likely to have an output level much greater than that of other observations with similar input levels. These outliers are more likely to have a super-efficiency score greater than 1. According to BANKER & CHANG [2006], DMUs with efficiency scores larger than 1.2 should be considered outliers and removed from the sample before conducting the final DEA calculation.

4 Model Specification and Data

4.1. Specification

In our empirical DEA model, R&D investments and manpower serve as inputs while patent applications are used to approximate innovative output. Some authors [e.g., ROUSSEAU & ROUSSEAU, 1997, 1998] suggest including publications as an additional output; however, we do not, for three reasons. First, recent studies reveal a number of measurement problems inherent in publication counts, such as double-counting in the case of co-authoring [SHARMA & THOMAS, 2008]. Second, detailed publication data are not available at the industry level; therefore, assigning publications to industries is highly problematic and would involve the difficult and possibly not entirely objective task of matching journals to sectors. Third, publication counts have the potential to introduce a language bias in favor of Anglophone countries.

We estimate a cross-industry cross-country pooled frontier, where each observation is accounted for as a single industry-country combination in time without considering the panel structure of the data. Since the objective of business R&D is to increase innovative output so as to improve the firm’s competitive position, we apply an output-oriented DEA model.

4.2 Data

This study analyzes research efficiency based on a sample of 13 EU member states and four nonmember states (Australia, Japan, South Korea, and the United States). The data on input and output for the efficiency analysis are collected from two underlying datasets: EU KLEMS and PATSTAT.

Our information on patent applications comes from the European Patent Office's Worldwide Patent Statistical Database.[3] This database, maintained by the European Patent Office, contains all national and international patent applications, including information on inventors and applicants and the location of each, priority dates, and technological classifications. We focus on EPO applications since an application to an international authority, in contrast to one made to a national authority, can be taken as a signal that the patentee believes the invention to be of high enough value to justify the expense of in international application. Central to our exercise is the construction of patent aggregates by country, industry, and year. We build this variable by using all patent applications filed with the European Patent Office (EPO) with a priority date between 2000 and 2004. The "priority date" is the date the invention was covered by a patent for the first time. However, most patents are first filed for at the national level and thus the majority of patent applications at the EPO are second filings. Accordingly, in this study, we date patent applications using the priority date instead of the usual application date since it is the date closest to the date of invention and the decision to seek patent protection [DE RASSENFOSSE & VAN POTTELSBERGHE DE LA POTTERIE, 2007]. Patent applications are assigned to the inventor's country, instead of that of the applicant, as the former is more indicative of the location where the invention occurred. In line with previous literature, only the first inventor's country of residence is considered [e.g., WANG, 2007; WIPO, 2008].

Patents are assigned to industries based on a concordance developed by SCHMOCH & AL. [2003], who used expert assessments and micro-data evidence on the patent activity of firms in the manufacturing industry[4] to link technologies to industry sectors. The international patent classification (IPC) technology classes provided in the patent application are grouped into 44 technological fields and then assigned to industries based on the NACE[5] code. Because patent applications usually contain more than one technology class and none of them

3 PATSTAT 1/2008

4 The authors argue that patents are most widely used in the manufacturing sector to protect intellectual property.

5 Nomenclature générale des activités économiques dans les Communautés européennes.

can be interpreted as its main class,[6] a weighting scheme is needed so as to avoid double counting of patents. Therefore, we weight every technology class mentioned in an application by the reciprocal of the total number of classes when constructing our industry-specific patent aggregates at the country level, which serves as the output in our efficiency analysis.[7] Finally, some further aggregation of NACE classes is needed to match the patent data to the input data sources. A detailed description of the concordance is provided in Appendix A.

Human capital and R&D effort serve as the inputs in our model. The R&D resources used in the innovative process at the sector level are approximated by R&D stocks provided by the EU KLEMS[8] database. From a theoretical point of view, R&D stocks are preferable to annual R&D expenditures since they capture the amount of knowledge available in an economy even though, in practice, assumptions must be made when calculating the initial stock. R&D stocks in the EU KLEMS database are built according to the perpetual inventory method,[9] as suggested by GUELLEC & VAN POTTELSBERGHE DE LA POTTERIE [2001]. To ensure comparability at the country level, the R&D stocks are deflated using implicit purchasing power parities[10] from the OECD [2008b] Main Science and Technology database.

The manpower invested in R&D is usually captured by the number of researchers per country as published by the OECD in the Main Science and Technology Indicators [OECD, 2008b]. However, these data are not available at the sector level and so we approximate human capital input by the share of skilled workers as we are convinced that researchers and support staff are mainly recruited from this group. The exact distinction between high-skilled and medium-skilled workers is of necessity vague due to differences in national educational systems [TIMMER & AL. 2007, 2008]. In case of high-skilled labor, comparability can be assumed for bachelor degrees, but not for any others. Therefore, we decided to include both high- and medium-skilled labor as inputs to control for heterogeneity across countries. However, our findings suggest that the main results are robust with respect to the use of skilled or high-skilled labor. Data on the share of high- and medium-skilled labor at the sector level are available from the EU KLEMS database. These shares are used to derive the amount of high- and medium-skilled labor in each industry and serve as additional inputs in the analysis of

6 This is in contrast to applications made at the United States Patent & Trademark Office (USPTO), which distinguishes between main and subclasses.

7 An example would be a patent with five IPC classes, each contributing only a fifth to the country-industry level aggregates.

8 Release March 2008.

9 The depreciation rate equals 12%. Calculation of R&D stocks is explained in detail in O'Mahony & al. [2008].

10 We use PPPs at constant 2000 prices, which are derived from R&D expenditures.

research efficiency.

Our dataset covers 13 industries for the period 2000 to 2004.[11] We impose one restriction on the industry-specific country patent aggregates, namely, that at least 15 patents were applied for within a certain year, to make sure that sufficient patent activity is present in each sector of the countries covered.

Table 1 sets out sample statistics of the input and output variables used in our analysis for the period 2000-2004. On average, across countries, industries, and years, 886 patents have been applied for at the EPO, although there is a great deal of heterogeneity within this average, ranging from a minimum of 16 patents to a maximum of 17,664. A similar pattern can be seen in the R&D stocks, measured in purchasing power parities to the basis year 2000. In line with expectations, the share of high-skilled workers is substantially smaller (one-fourth) than the share of medium-skilled workers.

Output Variable	**Observations**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
EPO patents	816	885.712	2266.49	16	17664
Input Variable					
R&D Stock (PPP)	653	12479.4	40855.95	1.13	370589.2
High Skilled	846	107.4	232.21	0.11	2008.9
Medium Skilled	846	428.76	583.66	0.74	3355.31

Table 1: Descriptive statistics

Looking at the country-level statistics, namely, the aggregated manufacturing-level data (Appendix C), we find that the United States has the highest average number of patent applications at the EPO, which is remarkable considering the “home” bias of the European countries in our sample. In Europe, Germany is the most frequent patent applicant, with an average R&D stock almost twice that of France. Comparing the number of high-skilled and medium-skilled workers, we find substantial variation across countries. Notably, the number of high-skilled workers in South Korea is more than four times that of Germany.

Appendix B shows the industry-specific means of the input and output variables calculated by averaging over countries and years. The two industries in our sample that exhibit the highest

11 The truncation point is determined by the availability of patent applications, which are published 18 months after application.

patent intensity are chemicals and chemical products and electrical and optical equipment. Both industries also have comparatively high R&D stock.

Consistent with recent literature on research efficiency [SHARMA & THOMAS, 2008; WANG & HUANG, 2007], we impose a lag structure for inputs to account for the fact that R&D efforts do not immediately result in innovative output [HALL & AL., 1986]. Therefore, inputs are lagged by two years in the DEA application.

5 Results

The empirical analysis is divided into three parts. First, we derive efficiency estimates for the manufacturing sector at the country level. Second, we identify industries occupying the world technology frontier by proceeding to industry- and country-specific data. Third, we focus on those industries revealing the highest efficiency scores — thereby defining the frontier — and conduct separate efficiency analyses for the industries of interest.

5.1 Cross-country comparison

A first impression of research efficiency in manufacturing is given by comparing average efficiencies at the country level. Figure 1 displays these average efficiencies for the period from 2000 to 2004. Averages are derived by first aggregating over sector-level data and then conducting a variable returns to scale[12] DEA analysis using these country-level aggregates. We implicitly assume of a time-invariant technology frontier and focus on the distance of countries from the estimated frontier. An alternative method would be to compare the technology frontiers of different years by means of Malmquist indices, as suggested in COELLI & AL. [2005].[13]

We find that Germany and Denmark are the most efficient countries with respect to research output in manufacturing, followed by the United States, the Netherlands, and Belgium (Figure 1). These countries could serve as benchmarks to help less efficient countries improve their performance. The high average efficiency of the United States, indicative of a remarkably strong position in the international context, is especially noteworthy due to our use of

12 As shown by Sharma & Thomas [2008], most countries reveal increasing returns to scale, hence, a constant returns to scale technology is inappropriate.

13 This approach is impossible in case of unbalanced panels and therefore not applicable for our datasets because we do not observe sufficient patent activity across all years, countries, and sectors.

European patent data to approximate innovative output.

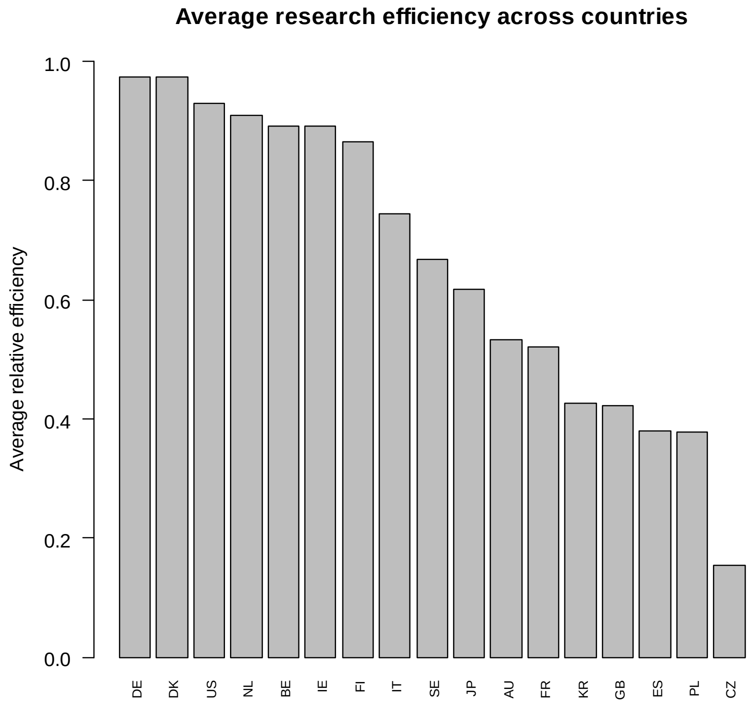

Figure 1 Average research efficiency across countries

This approach of using European patent data will tend to underestimate the output and thus the performance of non-European countries such as the United States, Japan, Australia, and South Korea. Inventors in these countries will first seek patent protection in their home markets and expand protection globally only for the most potentially profitable inventions. Thus, the United States is one of the leading countries worldwide in research and development in manufacturing. The leading role in Europe is played by Germany, which is located on or close to the technology frontier for all the years in our sample, thus revealing its excellence in research. Our results for the United States and Germany confirm those found by previous work [CULLMANN & AL. 2009; LEE & PARK, 2005].

Our results for total manufacturing are summarized by sorting our sample countries into three groups according to their average research efficiency in manufacturing:

- *high efficiency*: Germany, Denmark, the United States, the Netherlands, Belgium, Ireland, Finland;
- *medium efficiency*: Italy, Sweden, Japan, Australia, France;

- *low efficiency*: South Korea, the United Kingdom, Spain, Poland, the Czech Republic.

The small European economies, Denmark, Belgium, the Netherlands, Ireland, and Finland, show a remarkably high level of research efficiency, whereas some of the larger ones, namely, the United Kingdom, France, and Spain, lag behind. One explanation for this could be that it may be easier for small countries to link research conducted at universities to private business R&D activities due to the small number of large companies in those countries. Furthermore, a small country tends to show a higher degree of specialization, which could raise efficiency in the industries observed here.

The efficiency values for South Korea and Poland should be interpreted with caution because fewer data are available for these countries, especially R&D data at the sector level. Our results suggest that South Korea is not yet a major player in international innovation, but this could change in the near future because recent data show a drastic increase in Korean patent activity, both locally and at the international level [OECD, 2008a].

The lowest efficiency score was found for the Czech Republic, which is only slowly entering the international patenting arena. Recently, however, the country has increased its R&D efforts, and a 2008 OECD publication [OECD, 2008a] reveals that the Czech Republic is engaging in a great deal of cooperation with foreign co-inventors. Thus, our first-inventor approach to determining an invention's country of origin might contain a downward bias in the case of the Czech Republic, as the domestic inventor is often named second in international patent applications.

5.2 Analysis across countries and industries

The next step in our empirical analysis is to measure research efficiency across countries and industries by conducting DEA using a pooled sample of industry-country observations.[14] Therefore, we test whether the underlying technology exhibits constant or variable returns to scale. A p-value of 7.7 percent for the SIMAR & WILSON [2002] test statistic suggests rejecting the hypothesis of constant returns to scale. Hence, we allow for variable returns to scale in frontier estimation.

The assumption of a constant technology frontier enveloping all industries will be relaxed in

14 Poland and the Czech Republic have to be dropped due to insufficient data at the sector level.

the next section when we carry out an industry-specific efficiency analyses. To ensure the estimation of a consistent and robust technology frontier across countries and industries, we apply ex ante outlier detection by means of super-efficiency analysis [BANKER & CHANG, 2006].

A first impression of research efficiency at the industry level is achieved by comparing the average scores across industries. Therefore, we derive pooled cross-section frontier estimates where each observation is accounting for one industry in a certain country in one year and then average over countries, as shown in Table 2.

Industry	**Average efficiency**
Food products, beverages, and tobacco	0.114
Textiles, textile products, leather, and footwear	0.232
Wood, products of wood and cork	0.250
Pulp, paper, paper products, printing, and publishing	0.175
Coke, refined petroleum products, and nuclear fuel	0.219
Chemicals and chemical products	0.531
Rubber and plastics products	0.542
Other nonmetallic mineral products	0.505
Basic metals and fabricated metal products	0.299
Machinery, NEC	0.591
Electrical and optical equipment	0.638
Transport equipment	0.216
Manufacturing, NEC; recycling	0.454

Table 2 Average research efficiency at the industry level

The intertemporal frontier estimation exhibits average technical efficiencies of between 0.11 and 0.64, which are relatively low compared to those found in other empirical work. These results suggest that large inefficiencies are present within the knowledge production process. The low mean efficiencies are influenced by the large within-sample variation in research efficiency across countries.

There are substantial differences in patent intensity across industries. Chemicals, pharmaceuticals, and information and communication technology are known to be among the most patent-intensive industries, followed by machinery [SHEEHAN & AL., 2004]. This phenomenon may be due to different strategic motives for patenting in these industries, leading to patent fences or patent thickets, both of which accelerate patenting [NOEL & SCHANKERMAN, 2006; SCHNEIDER, 2008]. Therefore, one could argue that it is not surprising to find a higher average efficiency in electrical and optical equipment, chemicals (including pharmaceuticals), plastics products, and machinery simply because these industries tend to seek patent protection more often than do other sectors. However, these industries also exhibit greater R&D intensity and larger R&D stocks compared to others, as shown in our descriptive statistics in Section 4. Hence, our results suggest that the observable knowledge production process is more efficient in these industries and thus defines the research technology frontier. Table 3 lists the efficient combinations that suggest excellent research performance.

Industries	**Countries**
Wood, products of wood and cork	Italy
Coke, refined petroleum products, and nuclear fuel	Netherlands
Chemicals and chemical products	Germany (3)
Rubber and plastics products	Netherlands, Finland
Other nonmetallic mineral products	Denmark (3), Finland (2), Italy
Machinery, NEC	Italy (3), Germany, Netherlands
Electrical and optical equipment	Netherlands (2), Germany, United States, Finland
Transport equipment	Denmark
Manufacturing, NEC; recycling	Germany, Sweden, Italy

Table 3 Efficient country-industry combinations; number in parentheses indicates number of years country has occupied the technology frontier in the particular industry

The electrical and optical equipment industry is efficient in the Netherlands, Germany, the United States, and Finland. Due to the underlying panel structure of our data, we usually observe industries in countries for five consecutive years. However, a certain country-industry combination does not necessarily have to be efficient every year to stay at the technology frontier and that is exactly what we observe: country-industry combinations occupy the frontier for one or two years and lag slightly behind for the rest of the estimation period. An example is the German electrical and optical equipment industry, which is fully efficient only once but reaches an average efficiency of 0.93. This is the second highest value in the cross-country comparison; only the United States outperforms Germany, with an average of 0.96 in the electrical and optical equipment industry. Hence, the high research efficiency in this industry is one of the driving forces behind the high overall U.S. efficiency score.

Other industries that stand at the technology frontier include machinery, rubber and plastics, and chemical products. Chemicals and chemical products encompass the pharmaceutical industry, where patent protection has very strong effects because the process of research and development is so costly and time consuming that firms need to ensure protection of their intellectual property by way of a temporary monopoly [COHEN & AL., 2000]. Germany's chemical industry reaches the frontier in three out of five years, which emphasizes Germany's leading position, and not only in this industry; it also has large average efficiency scores of 0.93 and 0.89 for machinery and rubber and plastics, respectively. Our results confirm that the small European countries, Finland, the Netherlands and Denmark, are some of the best-performing countries in terms of research efficiency, with special strength in certain industries: Finland shows an excellent performance in rubber and plastics and mineral products, while Denmark plays a leading role in transport equipment. The Netherlands actually reaches the frontier in four industries, including machinery and electrical and optical equipment. Overall, we find electrical and optical equipment to be the most important industry when determining the technology frontier, followed by machinery.

5.3 Industry-specific analysis

We now relax the assumption of a common technology frontier and conduct separate industry-specific frontier estimations to identify leading countries, as well as those lagging behind, for the main industries of interest: electrical and optical equipment, machinery, and chemical products. The economic importance of these industries in the countries can be seen from Table 4, which sets out each industry's share of a country's gross output in total manufacturing.

Country	Chemical Products	Machinery	Electrical & Optical Equip.	∑
Australia	7.61%	5.48%	3.25%	16.34%
Belgium	16.74%	4.78%	5.13%	26.64%
Denmark	10.90%	12.52%	11.51%	34.93%
Finland	6.39%	11.63%	19.51%	37.54%
France	11.64%	6.92%	9.54%	28.09%
Germany	9.51%	12.58%	12.74%	34.83%
Ireland	26.83%	1.64%	28.69%	57.16%
Italy	8.24%	12.30%	8.21%	28.75%
Japan	9.22%	8.92%	16.92%	35.05%
Netherlands	18.41%	7.67%	8.31%	34.39%
South Korea	10.76%	7.04%	22.34%	40.14%
Spain	8.46%	5.51%	5.78%	19.76%
Sweden	8.51%	11.24%	12.55%	32.30%
United Kingdom	11.29%	7.50%	10.08%	28.87%
United States	11.03%	7.12%	13.45%	31.60%

Table 4 Share in total manufacturing of gross output

Running separate DEA analysis for the frontier industries generally corroborates our earlier findings. Germany and Denmark occupy the research frontier along with the United States and the Netherlands. We observe a relatively weak performance on the part of South Korea, the United Kingdom, and Spain, indicating that these countries have the potential to raise output given their levels of R&D effort. Once again, the score for South Korea should be interpreted with caution.

In the case of electrical and optical equipment, Australia joins the group of leading countries, whereas the United Kingdom shows the weakest performance.

In regard to the machinery industry, our earlier results showed this sector as efficient in Italy, Germany, and the Netherlands. Italy’s proficiency in this sector is confirmed by the present estimation results. The group of highly efficient countries in machinery also includes Belgium and Ireland. However, all the other countries exhibit a sharp decline in research efficiency, with Japan, Spain, the United Kingdom, and the United States all occupying surprisingly weak positions.

Country	Chemical Products	Machinery	Electrical & Optical Equip.
Australia	0.95	0.53	0.72
Belgium	0.77	0.94	0.81
Denmark	0.97	0.91	0.92
Finland	0.86	0.59	0.82
France	0.87	0.62	0.70
Germany	0.99	0.93	0.94
Ireland	0.72	0.96	0.56
Italy	0.77	0.99	0.40
Japan	0.52	0.36	0.83
Netherlands	1.00	0.94	0.81
South Korea	0.47	0.53	0.50
Spain	0.52	0.34	0.28
Sweden	0.54	0.52	0.56
United Kingdom	0.35	0.34	0.55
United States	0.99	0.44	0.96

Table 5 Efficiency scores for various industries

In the chemicals and chemical products industry, Germany is again the dominant player. The industry-specific analysis confirms the already identified leading groups of countries, with Australia close behind. At the end of the distribution, we find the United Kingdom, South Korea, Spain, and Japan, with a low average efficiency of about 0.5. Even though Japan is known for is pharmaceutical industry, the patent activity covered by the EPO dataset reveals substantial inefficiencies in the process of research and development, even when accounting for the home bias in patent applications.

This DEA application's focus on a single industry has given us a clearer picture of the strengths and weaknesses of our countries and, more specifically, of the gap between those that are efficient and those that are not. Compared to other industries, the efficiency gap in machinery production most obviously separates the countries into two groups: highly efficient and barely efficient.

6 Conclusions

This paper analyzes research efficiency at the industry level in total manufacturing for 13 European member and four nonmember countries between 2000 and 2004. We consider three inputs: knowledge stocks approximated by R&D expenditures and high- and medium-skilled labor to capture manpower.

The results on overall manufacturing can be summarized by sorting the countries into three groups according to their average research efficiency score:

- *high efficiency*: Germany, Denmark, the United States, the Netherlands, Belgium, Ireland, Finland;
- *medium efficiency*: Italy, Sweden, Japan, Australia, France;
- *low efficiency*: South Korea, the United Kingdom, Spain, Poland, and the Czech Republic.

The smaller European economies, namely, Denmark, Belgium, the Netherlands, Ireland, and Finland, have remarkably high levels of research efficiency, whereas some of the larger ones — the United Kingdom, France, and Spain — lag behind.

At the industry level, we find electrical and optical equipment to be the most important industry when determining the technology frontier, followed by machinery. Running separate DEA analyses for selected industries further supports the findings from the pooled estimation. Furthermore, estimating distinct industry frontiers paints a clearer picture of national strengths and weaknesses and, more specifically, shows more clearly the size of the gap between efficient countries and those that are less so.

Our results can provide guidance to policymakers interested in improving innovative performance and thereby ensuring long-term economic growth. Specifically, in a case of limited resources, priority should be given to those industries promising the largest output for the available amount of investment. However, the findings of this study should not be inappropriately overgeneralized, but viewed more broadly as general, indeed somewhat theoretical, advice

An interesting avenue of exploration for future research would be to test the influence of industry structure variables (e.g., competition or concentration) on research efficiency. This could be done with the bootstrap procedure proposed by SIMAR &WILSON [2007], which permits valid inference in the second-stage truncated regression of the efficiency scores on environmental variables while showing that conventional approaches for drawing inference in truncated Tobit regressions are invalid.

7 References

ANDERSEN, P., PETERSEN, N. (1993), A procedure for ranking efficient units in data envelopment analysis, *Management Science*, 39:1261–1264.

BANKER, R. D., CHANG, H. (2006), The super-efficiency procedure for outlier identification, not for ranking efficient units, *European Journal of Operational Research*, 175:1311–1321.

BANKER, R. D., CHARNES, A., COOPER, W. W. (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30:1078–1092.

CHARNES, A., COOPER, W. W., RHODES, E. (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research*, 2:429–444.

COELLI, T., RAO, P., BATTESE, G. E. (2005), *An Introduction to Efficiency and Productivity Analysis*, Springer, New York.

COHEN, W. M., NELSON, R. R., WALSH, J. P. (2000), Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not), NBER Working Paper, no. 7552.

CULLMANN, A., SCHMIDT–EHMCKE, J., ZLOCZYSTI, P. (2009), *Innovation, R&D Efficiency and the Impact of the Regulatory Environment*—A *Two Stage Semi-Parametric DEA Approach*, DIW Discussion Paper, forthcoming.

DE RASSENFOSSE, G., VAN POTTELSBERGHE DE LA POTTERIE, B. (2007), Per un pugno di dollari: A first look at the price elasticity of patents, *Oxford Review of Economic Policy*, 23:588–604

GRILICHES, Z. (1979), Issues in assessing the contribution of R&D to productivity Growth, *Bell Journal of Economics*, 10:92–116.

GRILICHES, Z. (1990), Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, 28:1661–1707.

GUELLEC, D., VAN POTTELSBERGHE DE LA POTTERIE, B. (2001), *R&D and Productivity Growth*, OECD Science and Technology Working Papers, no. 3.

HALL, B. H., GRILICHES, Z., HAUSMAN, J. A. (1986), Patents and R&D: Is there a lag? *International Economic Review*, 27:265–283.

HALL, B. H., ZIEDONIS, R. H. (2001), The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995, *RAND Journal of Economics*, 32:101–128.

HAUSMAN, J. A., HALL, B. H., GRILICHES, Z. (1984), Econometric models for count data with an application to the patents—R&D relationship, *Econometrica*, 53:909–938.

JAFFE, A. (1986), Technology opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value, *American Economic Review*, 76:984–1001.

KORTUM, J. (1997), Research, patenting, and technological change, *Econometrica*, 65:1389–1419.

LEE, H.-Y., PARK, Y.-T. (2005), An international comparison of R&D efficiency: DEA approach, *Asian Journal of Technology Innovation*, 13:207–222.

NOEL, M., SCHANKERMAN, M. (2006), *Strategic Patenting and Software Innovation*, CEPR Discussion Paper, no. 5701.

OECD (2008a), Compendium of Patent Statistics, OECD, Paris.

OECD (2008b), *Main Science and Technology Indicators*, OECD, Paris.

O'MAHONY, M., CASTALDI, C., LOS, B., BARTELSMAN, E., MAIMAITI, Y., PENG, F. (2008), *EU KLEMS—Linked Data: Sources and Methods*, University of Birmingham, Birmingham.

PAKES, A. (1986), Patents as options: Some estimates of the value of holding european patent stocks, *Econometrica*, 54:755–784.

PAKES, A., GRILICHES, Z. (1984), Patents and R and D at the firm level: A first look. In: Z. GRILICHES (Ed.), *R&D, Patents, and Productivity*, University of Chicago Press, pp. 55–72.

PAKES, A., SCHANKERMAN, M. (1984), The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In: Z. Griliches (Ed.), *R&D, Patents, and Productivity*, University of Chicago Press, pp. 73–88.

ROUSSEAU, S., ROUSSEAU, R. (1997), Data analysis as a tool for constructing scientometric indicators, *Scientometrics*, 40:45–46.

ROUSSEAU, S., ROUSSEAU, R. (1998), The scientific wealth of European nations: Taking effectiveness into account, *Scientometrics*, 42:75–87.

SCHERER, F. M. (1965), Firm size, market structure, opportunity, and the output of patented inventions, *American Economic Review*, 55:1097–1125.

SCHMOCH, U., LAVILLE, F., PATEL, P., FRIETSCH, R. (2003), *Linking Technology Areas to Industrial Sectors*, Final report to the European Commission, ISI, Karlsruhe.

SCHNEIDER, C. (2008), Fences and competition in patent races, *International Journal of Industrial Organization*, 26:1348–1364.

SHARMA, S., THOMAS, V. J. (2008), Inter-country R&D efficiency analysis: An application of data envelopment analysis, *Scientometrics*, 76:483–501.

SHEEHAN, J., MARTINEZ, C., GUELLEC, D. (2004), Understanding business patenting and licensing: Results of a survey. In: OECD (Ed.), *Patents, Innovation and Economic Performance: Conference Proceedings*, pp. 89–110.

SIMAR, L., WILSON, P. (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science*, 44:49–61.

SIMAR, L., WILSON, P. (2000), A general methodology for bootstrapping in nonparametric

frontier models, *Journal of Applied Statistics*, 27:779–802.
SIMAR, L., WILSON, P. (2002), Non-parametric tests of returns to scale, *European Journal of Operational Research*, 139:115–132.

SIMAR, L., WILSON, P. (2007), Estimation and inferences in two-stage, semi-parametric models of production process, *Journal of Econometrics*, 136:31–64.

TIMMER, M., O'MAHONY, M., VAN ARK, B. (2008), *The EU KLEMS Growth and Productivity Accounts: An Overview*, University of Groningen & University of Birmingham, Groningen and Birmingham.

TIMMER, M., VAN MOERGASTEL, T., STUIVENWOLD, E., YPMA, G., O'MAHONY, M., KANGASNIEMI, M. (2007), *EU KLEMS Growth and Productivity Accounts, Version 1.0, Part I Methodology*, University of Groningen, Groningen.

WANG, E. C. (2007), R&D efficiency and economic performance: A cross-country analysis using the stochastic frontier approach, *Journal of Policy Modelling*, 29:345–360.

WANG, E. C., HUANG, W. (2007), Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach, *Research Policy*, 36:260–273.

WIPO (2008), *World Patent Report—A Statistical Review*, WIPO, Geneva.

Appendix A: Concordance assigning IPC-Classes to European NACE[15]

NACE Revision 1	Industry	IPC classes
15t16	Food products, beverages, and tobacco	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, A23P, C12C, C12F, C12G, C12H, C12J, C13F, C13J, C13K, A24B, A24D, A24F
17t19	Textiles, textile products, leather, and footwear	D04D, D04G, D04H, D06C, D06J, D06M, D06N, D06P, D06Q, A41B, A41C, A41D, A41F, A43B, A43C, B68B, B68C
20	Wood, products of wood and cork	B27D, B27H, B27M, B27N, E04G
21t22	Pulp, paper, paper products, printing, and publishing	B41M, B42D, B42F, B44F, D21C, D21H, D21J
23	Coke, refined petroleum products, and nuclear fuel	C10G, C10L, G01V
24	Chemicals and chemical products	B01J, B09B, B09C, B29B, C01B, C01C, C01D, C01, C01G, C02F, C05B, C05C, C05D, C05F, C05G, C07B, C07C, C07F, C07G, C08B, C08C, C08F, C08, C08J, C08K, C08L, C09B, C09C, C09D, C09K, C10B, C10C, C10H, C10J, C10K, C12S, C25B, F17C, F17D, F25J, G21F, A01N, B27K, A61K, A61P, C07D, C07H, C07J, C07K, C12N, C12P, C12Q, C09F, C11D, D06L, A62D, C06B, C06C, C06D, C08H, C09G, C09H, C09J, C10M, C11B, C11C, C14C, C23F, C23G, D01C, F42B, F42D, G03C, D01F
25	Rubber and plastics products	A45C, B29C, B29D, B60C, B65D, B67D, E02B, F16L, H02G
26	Other nonmetallic mineral products	B24D, B28B, B28C, B32B, C03B, C03C, C04B, E04B, E04C, E04, E04F, G21B
27t28	Basic metals and fabricated metal products	B21C, B21G, B22D, C21B, C21C, C21D, C22B, C22C, C22F, C25C, C25F, C30B, D07B, E03F, E04H, F27D, H01B, A01L, A44B, A47H, A47K, B21K, B21L, B22F, B25B, B25C, B25F, B25G, B25H, B26B, B27G, B44C, B65F, B82B, C23D, C25D, E01D, E01F, E02C, E03B, E03C, E03D, E05B, E05C, E05D, E05F, E05G, E06B, F01K, F15D, F16B, F16P, F16S, F16T, F17B, F22B, F22G, F24J, G21H
29	Machinery, NEC	B23F, F01B, F01C, F01D, F03B, F03C, F03D, F03G, F04B, F04C, F04D, F15B, F16C, F16D, F16F, F16H, F16K, F16M, F23R, A62C, B01D, B04C, B05B, B61B, B65G, B66B, B66C, B66D, B66F, C10F, C12L, F16G, F22D, F23B, F23C, F23D, F23G, F23H, F23J, F23K, F23L, F23M, F24F, F24H, F25B, F27B, F28B, F28C, F28D, F28F, F28G, G01G, H05F, A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01M, B27L, B21D, B21F, B21H, B21J, B23B, B23C, B23D, B23G, B23H, B23K, B23P, B23Q, B24B, B24C, B25D, B25J, B26F, B27B, B27C, B27F, B27J, B28D, B30B, E21C, A21C, A22B, A22C, A23N, A24C, A41H, A42C, A43D, B01F, B02B,

15 Based on Schmoch & al. [2003].

		B02C, B03B, B03C, B03D, B05C, B05D, B06B, B07B, B07C, B08B, B21B, B22C, B26D, B31B, B31C, B31D, B31F, B41B, B41C, B41, B41F, B41G, B41L, B41N, B42B, B42C, B44B, B65B, B65C, B65H, B67B, B67C, B68F, C13C, C13D, C13G, C13H, C14B, C23C, D01B, D01D, D01G, D01H, D02G, D02H, D02J, D03C, D03D, D03J, D04B, D04C, D05B, D05C, D06B, D06G, D06H, D21B, D21D, D21F, D21G, E01C, E02D, E02F, E21B, E21D, E21F, F04F, F16N, F26B, H05H, B63G, F41A, F41B, F41C, F41F, F41G, F41H, F41J, F42C, G21J, A21B, A45D, A47G, A47J, A47L, B01B, D06F, E06C, F23N, F24B, F24C, F24D, F25C, F25D, H05B
30t33	Electrical and optical equipment	B41J, B41K, B43M, G02F, G03G, G05F, G06C, G06D, G06E, G06F, G06G, G06J, G06K, G06M, G06N, G06T, G07B, G07C, G07D, G07F, G07G, G09D, G09G, G10L, G11B, H03K, H03L, H02K, H02N, H02P, H01H, H01R, H02B, H01M, F21H, F21K, F21L, F21M, F21S, F21V, H01K, B60M, B61L, F21P, F21Q, G08B, G08G, G10K, G21C, G21D, H01T, H02H, H02M, H05C, B81B, B81C, G11C, H01C, H01F, H01G, H01J, H01L, G09B, G09C, H01P, H01Q, H01S, H02J, H03B, H03C, H03D, H03F, H03G, H03H, H03M, H04B, H04J, H04K, H04L, H04M, H04Q, H05K, G03H, H03J, H04H, H04N, H04R, H04S, A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, A62B, B01L, B04B, C12M, G01T, G21G, G21K, H05G, F15C, G01B, G01C, G01D, G01F, G01H, G01J, G01M, G01N, G01R, G01S, G01W, G12B, G01K, G01L, G05B, G08C, G02B, G02C, G03B, G03D, G03F, G09F, G04B, G04C, G04D, G04F, G04G
34t35	Transport equipment	B60B, B60D, B60G, B60H, B60J, B60, B60L, B60N, B60P, B60Q, B60R, B60S, B60T, B62D, E01H F01L, F01M, F01N, F01P, F02B, F02D, F02F, F02G, F02M, F02N, F02P, F16J, G01P, G05D, G05G, B60F, B60V, B61C, B61D, B61F, B61G, B61H, B61J, B61K, B62C, B62H, B62J, B62K, B62L, B62M, B63B, B63C, B63H, B63J, B64B, B64C, B64D, B64F, B64G, E01B, F02C, F02K, F03H
36t37	Manufacturing, NEC; recycling	A41G, A42B, A44C, A45B, A45F, A46B, A46D, A47B, A47C, A47D, A47F, A63B, A63C, A63D, A63F, A63G, A63H, A63J, A63K, B43K, B43L, B44D, B62B, B68G, C06F, F23Q, G10B, G10C, G10D, G10F, G10G, G10H

Appendix B: Summary statistics at the industry level

Industry	EPO Patents				R&D Stock				High Skilled				Medium Skilled			
	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum
Food products, beverages, and tobacco	1728.4	44.4	1688.0	1781.0	39514.3	2113.7	37416.1	41892.0	1466.1	45.6	1414.7	1539.0	8342.7	105.6	8240.3	8509.2
Textiles, textile products, leather, and footwear	1159.2	132.2	968.0	1311.0	10461.8	247.9	10109.2	10670.9	492.6	53.2	421.1	550.1	4409.0	502.6	3823.1	5084.7
Wood, products of wood and cork	178.4	40.5	139.0	229.0	3206.6	925.8	1838.5	3865.2	213.6	49.1	145.2	272.4	1888.6	265.7	1530.6	2236.2
Pulp, paper products, printing, and publishing	1211.8	81.1	1109.0	1290.0	24509.5	1870.3	22443.2	26875.3	2027.1	148.8	1836.7	2173.2	6162.9	306.4	5844.7	6595.8
Coke, refined petroleum products, and nuclear fuel	683.8	23.3	667.0	723.0	23958.7	1255.6	22586.2	25623.2	111.8	5.2	107.1	119.0	316.5	18.0	289.3	339.1
Chemicals and chemical products	28545.2	892.0	27214.0	29570.0	368141.9	13914.7	353882.7	384520.4	1586.9	30.7	1552.9	1628.1	3406.5	127.6	3261.5	3564.4
Rubber and plastics products	5617.4	106.9	5496.0	5734.0	37502.2	1949.3	35219.4	39634.4	928.7	31.4	893.7	962.9	4114.7	155.2	3984.6	4344.7
Other nonmetallic mineral products	3789.8	236.3	3487.0	4124.0	22539.5	245.4	22347.6	22865.9	526.8	9.0	518.7	538.9	2863.2	141.0	2713.2	3056.6
Basic metals and fabricated metal products	6307.6	128.1	6162.0	6455.0	62275.4	563.9	61605.1	62982.3	1798.8	46.2	1750.0	1869.9	10166.4	316.3	9891.5	10637.8
Machinery, NEC	24701.8	686.5	24001.0	25828.0	144652.2	6548.6	137205.5	151711.5	1911.0	109.7	1812.0	2066.9	7989.2	489.9	7545.3	8633.6
Electrical and optical equipment	56945.4	1828.0	55674.0	60165.0	779547.2	38031.6	735008.9	816686.9	4081.3	138.4	3916.6	4249.4	9973.3	899.6	9080.8	11099.8
Transport equipment	11288.0	802.7	10531.0	12345.0	502620.9	16632.0	488258.5	522170.6	2134.1	111.6	2049.3	2295.6	7145.6	157.9	7011.5	7367.8
Manufacturing, NEC	2256.8	65.4	2188.0	2343.0	15615.8	1050.5	14448.7	16803.1	695.6	22.8	669.9	721.5	3875.4	181.0	3721.8	4147.5

Appendix C: Summary statistics at the country level

Country	EPO Patents				R&D Stock				High Skilled				Medium Skilled			
	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum
Australia	988.73	396.14	120.00	1316.00	11475.31	731.60	10695.31	12383.88	233.67	25.12	187.79	265.04	912.41	23.81	873.58	942.25
Belgium	1791.46	513.58	703.00	2408.00	21247.66	691.09	20295.03	21869.15	100.09	2.44	96.31	104.06	500.09	27.46	451.92	534.94
Germany	31328.55	10153.03	6738.00	40494.00	235506.10	7169.71	227042.00	243748.10	923.38	27.03	899.79	984.07	7594.50	265.25	7175.89	8119.92
Denmark	1008.82	322.45	292.00	1377.00	8068.75	737.58	7192.70	8924.98	26.56	3.53	20.86	30.81	416.68	19.29	377.02	440.43
Spain	937.64	362.05	441.00	1631.00	16832.96	1105.64	15624.49	18158.09	523.04	119.31	318.33	685.37	1441.02	230.98	995.70	1730.36
Finland	1293.91	465.58	184.00	1756.00	12844.10	1342.51	11251.02	14380.25	185.62	16.26	155.60	205.53	346.51	19.14	311.06	369.24
France	8311.46	2627.65	1425.00	10909.00	126489.10	3404.08	122569.80	130383.50	417.88	34.32	379.91	505.69	3739.69	93.07	3543.70	3829.43
Great Britain	6117.46	1961.52	801.00	7673.00	97799.71	2146.41	95243.11	100233.10	781.08	69.20	660.10	851.81	5356.43	569.71	4325.81	6002.40
Ireland	214.45	89.01	47.00	329.00	3149.89	153.26	2965.63	3322.22	65.04	16.89	40.53	91.21	428.75	24.28	393.41	454.91
Italy	4929.91	1409.71	1909.00	6488.00	42905.19	114.67	42765.27	43019.69	248.24	14.71	225.13	266.75	8513.60	153.87	8127.31	8717.51
Japan	21125.64	7292.83	2606.00	27615.00	486848.40	18084.60	465781.70	507609.80	4215.53	74.92	4058.65	4320.42	15395.44	943.06	14044.39	16722.28
South Korea	1719.91	1323.35	526.00	4548.00	69024.85	3494.39	66553.94	71495.76	2750.78	430.54	2317.89	3472.43	6073.16	397.72	5340.16	6761.01
Netherlands	3431.82	1185.97	777.00	4747.00	24787.83	446.53	24222.00	25293.68	83.75	14.64	65.56	108.82	1319.26	55.68	1205.07	1375.66
Sweden	2441.73	634.41	728.00	3008.00	36348.87	2820.43	32826.70	39345.42	116.89	27.35	84.87	165.36	870.20	30.12	835.65	926.61
United States	33048.82	10558.20	3428.00	39608.00	880727.00	5370.19	873631.70	886484.20	7781.95	380.48	7083.66	8304.06	22283.02	2549.71	18570.38	24570.59