WHAT HAPPENS WHEN FIRMS PATENT?

NEW EVIDENCE FROM U.S. ECONOMIC CENSUS DATA

by

Natarajan Balasubramanian *
Florida International University

and

Jagadeesh Sivadasan *
University of Michigan

CES 08-03   January, 2008

Abstract

In this study, we present novel statistics on the patenting in US manufacturing and new evidence on the question of what happens when firms patent. We do so by creating a comprehensive firm-patent matched dataset that links the NBER patent data (covering the universe of patents) to firm data from the US Census Bureau (which covers the universe of all firms with paid employees). Our linked dataset covers more than 48,000 unique assignees (compared to about 4,100 assignees covered by the Compustat-NBER link), representing almost two-thirds of all non-individual, non-university, non-government assignees from 1975 to 1997. We use the data to present some basic but novel statistics on the role of patenting in US manufacturing, including strong evidence confirming the highly skewed nature of patenting activity. Next, we examine what happens when firms patent by looking at a large sample of first time patentees. We find that while there are significant cross-sectional differences in size and total factor productivity between patentee firms and non-patentee firms, changes in patent-ownership status within firms is associated with a contemporaneous and substantial increase in firm size, but little to no change in total factor productivity. This evidence suggests that patenting is associated with firm growth through new product innovations (firm scope) rather than through reduction in the cost of producing existing products (firm productivity). Consistent with this explanation, we find that when firms patent, there is a contemporaneous increase in the number of products that the firms produce. Estimates of (within-firm) elasticity of firm characteristics to patent stock confirm our results. Our findings are robust to alternative measures of size and productivity, and to various sample selection criteria.

Keywords: Innovation, productivity, new products, firm scope

JEL classification codes: O30, O31, O34, O33, L25

* The research presented here was conducted while Natarajan Balasubramanian was a Census Bureau research associate at the California Census Research Data Center. Research results and conclusions expressed are those of the authors, and do not necessarily indicate concurrence by the Bureau of the Census. The results presented here have been screened to ensure that no confidential data are revealed. We thank Scott Stern, Marvin Lieberman, Daniel Ackerberg, Hugo Hopenhayn, Michael Darby, Mariko Sakakibara, Arvids Ziedonis, Rosemary Ziedonis, Andrew Bernard and John Sutton for their comments and suggestions. We also thank seminar participants at the ASSA meetings in New Orleans, University of Michigan and University of California, Los Angeles. All remaining errors are our own.

1 Introduction

A number of central questions, in economics in general and industrial organization in particular, relate to differential innovativeness. Economists have long recognized that innovation is the driving force behind economic growth and hence crucial to improving living standards. Since a large amount of innovation is undertaken by firms, a very rich literature has emerged that tries to understand the micro-foundations of aggregate growth by examining firm-level innovation processes. Also, innovation has been recognized as crucial for long-term firm profitability, survival and growth, and hence been a subject of intense research in the industrial organization and management literature.

Patent statistics have been very widely used in studies of firm innovation at least since Scherer (1965). The primary motivator for using these data has been, as Griliches (1990) put it, "the dream of getting hold of an output indicator of inventive activity". In this context, Griliches poses two fundamental questions (i) what aspects of economic activity do patent statistics actually capture?, and (ii) what would we like them to measure? To answer the first and, as Griliches argues, more relevant question, it is important to understand whether and what (if any) real changes happen within firms when they patent. In this paper, we construct a comprehensive firm-patent data linkage and use it to address this question.

We start by using name-matching procedures to link the NBER patent data (constructed by Hall, Jaffe and Trajtenberg, 2001) to the business register (SSEL) data from the US Census Bureau. This new dataset significantly enhances coverage of patent data-firm data links; from about 4,100 unique patent assignees in the original patent data - Compustat data link (Hall, Jaffe and Trajtenberg, 2005), our linked dataset extends the coverage to more than 48,000 unique assignees, representing almost two-thirds of all non-individual, non-university, non-government assignees during 1975 to 1997. This linked dataset has a number of advantages over the existing Compustat link including coverage of non-listed firms, availability of detailed input and output data at establishment level and ability to track ownership changes over time.

Our comprehensive firm-patent dataset allows us to present some novel statistics on the prevalence and importance of patenting in the US manufacturing sector.[1] We find strong evidence for the highly skewed nature of patenting activity – just 5.5 percent of the firms

[1] While we linked the patent data to the business register covering all sectors, the rest of our study focuses on the manufacturing sector (as we had access to only the manufacturing sector census datasets for this study).

account for all of the patenting activity. These firms have a highly disproportionate share of economic activity, as they account for about 59.3 percent of value added, 63.5 percent of capital stock and about 52.2 percent of employment. We find significant heterogeneity across industries in the prevalence of patenting, broadly consistent with Scherer's (1983) findings using data on a sample of large US corporations. Examining temporal changes in the role of patentees, we document two interesting facts. First, between 1977 and 1997, aggregate output, value added and capital of patentees grew faster than non-patentees, pointing to an increasing economic role for patent-owning firms. Second, despite the increase in aggregate output, aggregate employment of patentees declined between 1977 and 1997 (by about 1.1 million). Interestingly, aggregate employment of non-patentees increased in the same time period. This suggests that improvement in labor productivity among patentees accounted for most of the decline (of about 1 million jobs between 1977 and 1997) in aggregate employment in the manufacturing sector.

We then turn to the fundamental question of our study - what happens when firms patent? To answer this question, we first examine cross-sectional differences between patent-owning and non-patent owning firms. We find that after conditioning out industry-year effects, patent-owning firms are much larger (by a factor of about 10), more skill intensive (by about 5 percent) and more capital intensive (by about 40 percent). We find that labor productivity is higher by about 23 percent, and TFP differences are about 15.2 percent (when using the Solow residual definition) and about 8 percent (when using the Ackerberg, Caves and Frazer (2005) methodology).

To understand if these cross-sectional differences simply reflect self-selection by larger and more productive firms into patenting, we identify about 9,200 firms that switched status from non-patentee to patentee during our panel period (i.e. firms that enter without a patent but apply for a patent sometime after entry). We then use the data on these "switchers", and present evidence on the timing of patenting effects.[2] To our knowledge, this is the first comprehensive study of first-time patentees. We find that applying for a patent is associated with a large and contemporaneous increase in the size of the firm, both in output (real revenue and value added) as well as in inputs (capital stock and employment). There is also some evidence of a contemporaneous increase in skill intensity, and correspondingly of an increase in labor productivity. Both capital intensity and total factor productivity

[2] In spirit, this approach is similar to that adopted by Clerides, Lach and Tybout (1998) for analyzing whether exporting increases productivity.

measures show some increase a few years before the patenting event, but very little change around the patenting event.

The weak evidence for increase in total factor productivity suggests that the large increase in size of firms following patenting may be driven by an increase in the scope of the firms. Consistent with this explanation, we find that patenting firms have a much larger number of products (measured as the distinct number of seven-digit product categories reported by the firm) relative to non-patentees. Also, examining data on switchers, we find that the number of products show a significant increase around the time of the first-time patenting event.

We also find similar results in regressions examining the elasticity of firm characteristics to patent stock. There is a statistically significant positive elasticity of firm size and scope (number of products) with respect to patent stock, but little evidence of any significant relationship between productivity (or capital and skill intensity) and patent stock. We find our results robust to alternative measures of size and productivity, and to various sample selection criteria. Finally, we find that post-patenting size increases are meaningfully associated with commonly used measures of underlying patent quality. Specifically, we find that the size increases are positively correlated with the number of patents filed in the first year as well as the number of forward citations and number of citations per patent for those patents.

Overall, our results strongly suggest that patenting is associated with real economic effects. With regard to the second question (of what we would ideally like patent statistics to measure), Griliches suggests we would like to use patent data to “measure and understand better the economic processes that lead to the reduction in the cost of producing existing products and the development of new products and services.” Our results suggest that at the firm level, patents are more likely to be associated with the latter (new product development) rather than the former (reduction in costs of existing products or increase in productivity). These findings are consistent with Levin et al. (1987), who found that firms they surveyed considered patents a more effective mechanism for protecting product rather than process innovations.

Our results also provide some guidance about the class of models to be used in the context of patenting by firms. Heterogeneous firm industry equilibrium models in the industrial organization literature could be classified broadly into two categories. The first class of models use productivity (lower marginal costs) as the source of heterogeneity, and hence size dispersion, across firms (e.g. Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes, 1995; Melitz, 2003). More recently, a second class of models (Klette and Kortum, 2004; Bernard Redding and Schott, 2006b; Nocke and Yeaple, 2006) use firm scope (number of products) as the source of firm heterogeneity. Our results suggest that the latter class may be more useful for modeling firm patenting behavior.[3]

This work builds on the tradition of a rich literature that has used firm-patent linked data to understand patenting and innovation at the firm level. Patents have been shown to be correlated with various important firm level outcomes such as market value (e.g. Griliches, 1981; Pakes, 1985; Austin, 1993; Bloom and Van-Reenen, 2002; Hall, Jaffe and Trajtenberg, 2005), productivity (e.g. Bloom and Van-Reenen, 2002) and access to capital (e.g. Kortum and Lerner, 2000; Mann and Sager, 2007).[4] To our knowledge, ours is the first study to closely analyze the timing of size and scope effects associated with patenting, using comprehensive firm-level data. This study (and many others in this literature) owes much to the efforts of Hall, Jaffe and Trajtenberg (2001), in putting together the NBER patent dataset and making it available for public use. We also take advantage of the efforts of researchers and staff at the US Census Bureau in maintaining and updating the business register.

This paper is organized as follows. The next section (Section 2) describes the methodology adopted to match the patent data to the SSEL, and presents an analysis of the match coverage. Section 3 provides descriptive statistics on the economic importance of patent-owning firms. Section 4 focuses on how patent-owning firms differ from non-patent owning ones on characteristics such as size, productivity and input choices. In Section 5 we analyze firms that switch status from non-patent owning to patent-owning firm, and examine changes in firm characteristics associated with the switch. In Section 6, we analyze the relationship between firm scope (number of products) and patenting status. In Section 7, we undertake a number of robustness checks of our results. In Section 8, we discuss our findings in the context of related empirical and theoretical literature. Section 9 concludes.

[3]Interestingly, the results on firm size, productivity and patenting that we document are similar to the empirical results on firm size, productivity and exporting documented by Bernard and Jensen (1999) and Clerides, Lach and Tybout (1998).

[4]Besides being a measure of the level of innovative activity within the firm, information contained in patents have been used in studies of other phenomena, such as the role of intellectual property rights regimes and appropriability (e.g. Sakakibara and Branstetter, 2001; Hall and Ziedonis, 2001; Branstetter, Fisman and Foley, 2006), technological diversity (e.g. Miller, 2006), and spillovers (e.g. Jaffe, Trajtenberg and Henderson, 1993).

2 The NBER-Census linked dataset

2.1 Matching NBER patent data to Census data

The patent assignee data in the NBER patent data were matched to the Census firm data using name-matching procedures. Since the Census data do not contain firm names, the names were obtained from the Standard Statistical Establishment List (SSEL). The resulting NBER Patent Data-SSEL Bridge links the assignee number to one or more Census Firm Numbers (CFN) for every year over the period 1975 to 1997. The SSEL is an annual dataset containing information such as ownership and industry code, and is linked to the Census firm data through the CFN. Using the SSEL to match the data also provides three distinct advantages. First, in contrast to the Census of Manufacturing that covers only manufacturing, the SSEL covers all industries. Though the eventual analyses and summary statistics were limited to manufacturing, this is important because the patent dataset does not contain any industry codes, and hence, assignees could potentially be matched to firms in non-manufacturing industries (i.e. industries other than SIC codes 2000 to 3999). Second, the SSEL covers all establishments in the US that pay payroll tax, and hence provides very extensive coverage of the US economy. Finally, the SSEL also tracks the ownership of each establishment every year through the CFN (establishments of a firm share the first few digits). While this link may be temporarily incorrect (usually by not more than a year) due to reasons such as mergers and acquisitions, it does provide a reasonably accurate picture of ownership over time. This is particularly important in tracking small patent assignees that begin as a single-unit firm, and eventually get acquired by bigger firms. The limitation of using the SSEL is that it does not cover establishments that do not have any employees (e.g. an entrepreneur in a proprietorship). Hence, any start-ups that are not incorporated and have no employees will not appear in the SSEL. However, at this time, there are no datasets that cover this segment of the population, and hence, any extension of this matching must wait till such a dataset is available.

Assignee names in the patent dataset were matched to names in the SSEL for each of the 23 years from 1975 to 1997, thus providing an assignee-year match to the SSEL (as opposed to a single assignee-CFN match across all years). This year-by-year match permits tracking of any ownership changes over time (albeit a bit imperfectly as noted above) unlike the existing Compustat match included in the NBER patent dataset which is based only on the 1989 corporate structure. Doing the matching year by year also considerably reduced the number of observations to be analyzed at any one time (from hundreds of millions to less than a tenth of that), thus expediting the execution of the name-matching algorithms.

An important restriction was that only US assignees (country code = "US") that were not individuals or governments (i.e. assignee code = 2) were selected for matching. The obvious reason for this restriction was that the SSEL did not cover individuals or foreign firms without establishments in the US. This restriction limits the patents to those belonging to firms, research institutes and universities. While it was not possible to completely eliminate these patents, universities were identified through a visual inspection of their names, and excluded after the matching was done.

The matching process for each year was done in two stages. The first step was obtaining a preliminary set of clusters using the SAS name-matching procedure DQMATCH. In addition to looking for names that are identical, this program also gathers into "clusters", names that are similar alphabetically as well as names that sound similar. Besides the names of the assignees (from the patent dataset) and the establishment (from the SSEL), the procedure also used geographic information available in these datasets. The patent dataset does not contain information on the city and state of the patent assignee; however, it does provide this information for all inventors on a patent. The SSEL contains the city and state for every establishment. These data were used to limit the name matching only to those establishments in the city and state of any of the assignee's inventors. In addition, a second set of clusters was obtained by using only information on the state (and ignoring the city) that was used to generate matches for assignees not matched in the first set of clusters.

While the DQMATCH procedure does a good job of combining similar names into clusters, these clusters usually contained many names, and hence, were not directly useful for creating the bridge. In the second step, the names of assignees and firms in these clusters were "cleaned" by applying certain rules to obtain more precise matches. Any assignee and an establishment of a firm whose "cleaned" names matched exactly were treated as a match, with the assignee number being "bridged" to the CFN of the establishment. A tiered approach to cleaning was used. First, the most restrictive rules were used to define a match. These rules were progressively relaxed to obtain additional matches, but potentially with a lower degree of confidence in the match. The resulting bridge had matches in five "reliability codes" depending on the geographic information used for match (city vs. state) and the rules applied for cleaning the names. A more detailed explanation including a flow-chart of the matching process is provided in Appendix 1.

2.2 Analysis of match coverage

Table 1 provides an overview of the coverage obtained in the NBER-SSEL bridge. The coverage is analyzed using two types of matches. The first, "ever-matched", refers to a patent assignee-SSEL firm match irrespective of the year of patenting or the year of occurrence in the SSEL. Hence, if an assignee applied for a patent only in 1981 (i.e. did not apply for patents at any other time), and was matched to an SSEL firm in the years 1984 and 1985, then, this would be classified as an "ever-match" even though the year of the application and the firm's appearance in the SSEL do not coincide. The advantage of this match is that timing errors such as those caused when a firm patents as a non-employee firm (and hence, is not included in the SSEL), and eventually has employees (and hence, appears in the SSEL) are addressed. However, the disadvantage is that it increases the potential for mismatches, and in addition, does not allow tracking of ownership changes. The second type of match is a "contemporaneous match", defined as a patent assignee-SSEL firm in the application year of patenting. Hence, in any given year, a contemporaneous match exists only if the assignee applied for a patent during that year, and can be found in the SSEL. When matched, this is clearly more accurate than the "ever-match", but the chances of a mismatch occurring due to timing errors such as the one discussed above are higher. While the coverage analyses are presented using both these types of matches, the contemporaneous match forms the basis for most of the subsequent analysis.

Table 1 presents the coverage statistics for various populations. Of the 2.92 million patents in the NBER Patent Dataset belonging to 175,115 assignees, about 40% belonging to about 57,600 assignees were matched at least once ("ever-match") to a firm in the SSEL. Within the relevant population of patents applied between 1975 and 1997 by US assignees that are not individual, universities or government agencies, 90% of patents, and 63.7% of assignees are "ever-matched". Using the contemporaneous match criterion, about 80.6% of patents, 79.7% of citation-weighted patents and 64% of assignee-years are matched. This provides a substantial increase in coverage over the existing Compustat-NBER match that covers about 4100 firms.

Figure 1 plots the year-by-year coverage for the contemporaneous match. Patent coverage hovers around 80% in most years with the exception of 1988 when it is 65%. This is because a large part of the SSEL firm data for this year was not available. When measured using citation-weighted patents, the annual coverage is similar, varying from 78% to 86% (with the exception of 1988). Assignee-level coverage varies from about 53% in 1988 to about 70% in 1977, with an average of about 64%. Hence, it appears that the contemporaneous match is reasonably uniform over time. The "ever-match" coverage (not presented here), expectedly, shows a mild declining trend over time from about 80% of assignees matched in the earlier years to about 70% in 1997. Table 2 presents the coverage by technological category. The NBER Patent Dataset provides 6 technological categories (based on over 400 US Patent Classes). While there is some variation across categories (about 75% of patents contemporaneously matched in Drugs and Medical to 84.5% in Electrical and Electronics), no technological category appears to be unreasonably under-matched or over-matched.

In order to identify any systematic biases in the matching process, Table 3 compares the matched and non-matched assignees in the NBER Patent Dataset. Not surprisingly, the matched assignees are larger in terms of number of patents (about 1.73 patents per year compared to 1.27 patents per year for the non-matches). They also appear to patent for longer periods of time as evidenced by the bigger difference when an "ever-matched" criterion is used (14.42 patents over 1975 to 1997 compared to 2.81 for the non-matches). While part of this difference is likely due to genuine reasons such as the smaller firms not being included in the SSEL, some of this difference could be due to inaccuracies in the matching process itself - e.g. a larger firm may have more establishments, and hence have a greater probability of being matched. Nevertheless, given that almost 80% of patents, and 64% of assignees were contemporaneously matched, we expect the errors due to non-matches to be relatively small. Finally, there appears to be almost no difference in the average quality of the patents between the matched and non-matched assignees. The average number of forward citations per patent is about 6.2 for both types of assignees, while the number of claims per patent is only slightly different (13.60 vs. 13.94).

The remainder of the paper is devoted to presenting summary statistics and analyzing various aspects of patenting in the US manufacturing sector. In order to develop these statistics, the NBER-SSEL Bridge discussed above was matched to Census data for the five Censuses between (and including) 1977 and 1997 (i.e. census years 1977, 1982, 1987, 1992 and 1997). Since Census data were at the establishment level, and patenting is a firm-level variable, the establishment data were aggregated to the firm level using the CFN. In case a firm had establishments in multiple SIC-4 industries, it was assigned to the largest SIC-4 industry as measured by share of firm output. Also, a firm could be matched to multiple assignees in a single year; hence, data from all assignees were used to compute patenting statistics for the firm. For instance, stock variables such as number of patents applied in the current year were computed as the sum of the number of patents applied by all assignees matched to the firm in that year. Indicator variables such as whether a firm was assigned a patent at any time till the current year were based on whether the firm was matched to any assignee till that point in time. This linked dataset contained about 1.4 million firm-year observations, and was used to develop a number of summary statistics, including the economic importance of patent-owning firms, differences in characteristics between patent-owning and other firms, and whether patenting is associated with any real changes in variables such as output and productivity, as discussed in the following sections.

3 Statistics on the role of patentees in manufacturing

In this Section, we use the matched dataset to document some basic statistics about the importance and extent of patenting in US manufacturing. As noted earlier, we constructed the cross-link between the patent data and the business register, which covers the universe of all employers (in all sectors) in the United States. However, for the rest of our analysis, we focus on the manufacturing sector. This is partly motivated by the fact that patenting is likely to be most common in sectors within manufacturing (Scherer, 1983), and partly by data access limitations.[5]

The first set of summary statistics in Table 4 highlight the economic importance of patent-owning firms in the US manufacturing sector. The first column in Table 4 gives the proportion of firms within a SIC-2 industry that own a patent at some point between 1963 and 1997. The most striking fact is that on average, only about 5.5% of all firms own patents. Furthermore, in no SIC-2 industry does this proportion exceed 20%. These figures are consistent with the stylized fact that a large fraction of firms report zero or little R&D expenditure (e.g. Cohen and Klepper, 1992), and with the findings in Cohen et al (2000) that patents are not viewed as very effective means of protecting profits from inventions in a majority of manufacturing industries. Not surprisingly, industries vary in the fraction of patent-owning firms - only about 0.5% of all firms in lumber and wood products (SIC 22) owned patents compared to a little over 17% in SIC 38 (Instruments and Related Products).

The next three columns of Table 4 present the share of patent-owning firms in industry value added, capital stock and employment. Even though they form a small fraction of firms

[5] For this project, we had access to the business register and US Census Bureau datasets on manufacturing. In future work, we propose to use our link to look closely at the prevalence of patenting in sectors outside of manufacturing.

by number, patent-owning firms appear to dominate economic activity. In as many as 11 out of the 19 industries, the small number of patent-owning firms account for more than half the industry value-added. On average, they account for more than 52% of all employment in manufacturing, and over 63% of all capital stock. Furthermore, even in industries where the share of patent-owning firms is very small (e.g. lumber and wood products), these firms have a significant economic role, as they contribute almost 30% of all value added. In industries such as transportation (SIC 37), chemicals (SIC 28) and instruments (SIC 38), almost nine-tenths of industry value added can be attributed to the relatively few firms that own patents. Another interesting aspect of Table 4 is that patent-owning firms generally account for a larger share of value added than of employment. This is true in every SIC-2 industry. Hence, on average, these firms seem to have much greater labor productivity.

It is well established that a major share of all R&D is concentrated in large firms (e.g. Cohen and Klepper, 1996). While confidentiality reasons prevent disclosure of detailed statistics on the proportion of patenting accounted for by large firms, the share of patenting firms in manufacturing activity by size decile is presented in Figure 2. As expected, the extent of patenting increases with size, but the increase is mostly concentrated in the largest size decile. Almost 80% of value added, and a little less than 70% of employment in the tenth size decile pertain to patent-owning firms. In contrast, in the ninth size decile (i.e. the second largest size decile), patent-owning firms account for less than 20% of value added and employment. However, the trends in the fraction of firms that own patents are similar across size deciles. Even in the largest size class, only about 20% of firms own patents, with the vast majority not being patent assignees.

Table 5 briefly examines the changes in the share of patent-owning firms over time. Column 1 provides the aggregate output, value added, capital stock and employment for patent-owning (i.e. firms that ever owned a patent) and non-patent-owning firms (i.e. firms that never owned a patent) in 1977. In line with trends in Table 4, patent owning firms accounted for about $1.46 trillion or about 68% of the total output of $2.14 trillion (in 1987 dollars). They also employed about 10.5 million of the 18.5 million employed in manufacturing. Columns 2 and 3 present the same statistics for 1987 and 1997, with the changes from 1977 to 1987, and from 1977 to 1997 included in Columns 4 and 5 respectively.

Two interesting facts stand out from this table. First, patent-owning firms have grown much faster than non-patent owning firms in terms of aggregate output, value added and capital stock. A part of this could be attributed to new entrants that come in with patents.

The second, and perhaps more interesting fact is that aggregate employment in patent-owning firms has actually declined over this same period. The reduction in employment among these firms is greater than the reduction of 1 million in all manufacturing employment. While this study does not explore these findings in detail, these two facts together imply that the labor productivity of patent-owning firms has increased considerably more than those of non-patent owning firms.

4 Overall differences between patentees and non-patentees

This section presents detailed statistics that shed light on how patent-owning firms differ from other firms. The statistics are classified into three broad categories - (i) size; (ii) skill and capital intensity; and (iii) productivity.

We use four measures of size – real output, real value added, capital stock and employment.[6] We measure skill intensity using two measures – the white-collar share of wages and the white- to blue-collar worker ratio. Capital intensity is measured as capital stock per worker. Productivity is measured in three ways. The first measure is labor productivity, which is defined as real output per worker. The second measure is total factor productivity, which is measured as a Solow residual (roughly defined as log output less cost share weighted sum of log inputs – see Appendix 2 for the precise definition). The Solow residual definition assumes that the production function is constant returns to scale and that input and output markets are perfectly competitive. To address potential biases in case these conditions do not hold, we use an alternative measure of TFP based on the methodology proposed by Ackerberg, Caves and Frazer (2005). In this methodology, productivity is defined as the residual from a regression of real output on real inputs, where a method of moments approach is used to correct for endogeneity of input choices caused due to unobserved productivity heterogeneity.[7]

All the variables, except for the skill intensity variables (which are ratios), are logged. This minimizes the influence of outliers in the distribution of these variables in levels, and allows for an easier interpretation of differences between patentees and non-patentees.

The first two columns of Table 6 provide the (unweighted) mean value of these firm characteristics for firms that have owned patents at any point till the current year, and for

[6] Refer appendix 2 for detailed definitions of all variables.

[7] Ackerberg, Caves and Frazer (2005) show that their approach addresses some of the shortcomings in the methodologies proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). A detailed discussion of the estimation algorithm is provided in Balasubramanian (2007).

other firms. The overall means in columns 1 and 2 may reflect industry specific differences, if patenting firms are more predominantly located in certain types of industries. To control for industry specific effects, we estimate the following regression:

$$\Phi_{ijt} = \beta_0 D_{it} + \eta_{jt} + \epsilon_{ijt}$$

where Φ_{ijt} is the variable of interest (e.g. output), D_{it} is 1 if the firm owns a patent, and 0 otherwise, η_{jt} are industry-year dummies. The estimated β_0 coefficients are presented in column 3. In our discussion of the results below, we largely focus on column 3.

First we examine differences in the size variables. Comparing the simple means of the size variables (columns 1 and 2), it is clear that on average, patent-owning firms are bigger by a factor of about 15 in terms of output, value added and capital stock, and are about 10 times bigger when employment is used as a measure of size.[8] The large size differences seen in columns 1 and 2 remain after industry-year effects are controlled for (Column 3). This results suggest that the size differences between patenting and non-patenting firms are substantial both economically and statistically, and are consistent with the findings in Table 4 that suggest an important economic role for these firms.

The next set of statistics in Table 6 examines differences in skill and capital intensity. Patent-owning firms exhibit higher skill-intensity (the white- to blue-collar worker ratio for these firms is higher by about 5.8%; the wage share of white-collar workers is higher by about 4.6%), and significantly greater use of capital (almost 35% higher capital per worker).[9]

Patent-owning firms have a higher level of productivity than other firms. The simple mean (log) output per employee is higher by about 0.4 log points (or a factor of 1.5) for patent-owning firms. Controlling for industry-year fixed effects, the difference is approximately 21.8%. TFP is also higher for patent-owning firms though the magnitude of the difference is not as high as that for labor productivity. This is consistent with patent-owning firms choosing higher levels of capital and more skilled (higher quality) workers than other firms. The difference is about 14% if the Solow residual is used, and about 8% when TFP is computed using the Ackerberg-Caves-Frazer method.

Taken together, Table 6 strongly suggests that patent-owning firms are quite different from their non-patenting counterparts. Specifically, they are much larger, tend to choose

[8] Note that these variables are logarithms of the original values, and hence the factor, say for output, is computed as $\frac{e^{8.915}}{e^{6.195}}$.

[9] In unreported results, we also examined differences in wage rates between patenting and non-patenting firms. We find that patentees pay higher wages (both for production and non-production workers). This is consistent with the use of higher quality/skilled workers by patenting firms.

higher levels of skilled labor and capital, and exhibit greater productivity. The next section goes beyond these pooled regressions, and attempts to understand what part of these differences are attributable to changes in patenting status within firms.

5 What happens when firms patent? Analysis of first-time patenting

The analyses in the last section covered all firms, irrespective of whether they entered the dataset with a patent or not. While that provides a good overview of the differences between patent-owning firms and other firms, it does not provide meaningful inferences on whether any part of these differences is due to the change in patenting status. For instance, it is possible that firms that eventually own patents are bigger, and more productive even as they enter, and patenting is simply an indicator of these initial differences. On the other hand, it could be the case that these firms were similar to other firms before they started patenting, and got bigger and more productive subsequent to patenting. In order to distinguish between these scenarios, this section focuses on "switchers" - firms that enter the dataset without a patent, and eventually own a patent - and examines whether the switch from being a non-patentee to a patentee firm is associated with significant changes in the firm's characteristics. To our knowledge, this is the first exercise of this kind in the patenting literature.

5.1 Regression analysis of first-time patenting

Table 7 presents OLS and within regressions of various firm characteristics for these "switchers" over their entire lifecycle, and in the ten-year period around the time of the "switch" (5 years prior and 5 years after the year of patenting). Specifically, the following regressions are estimated:

$$\Phi^M_{ijt} = \beta_0.D_{it} + \eta_i + \epsilon_{ijt}$$

where Φ^M_{ijt} is the variable of interest demeaned of industry-year fixed effects, and D_{it} is 1 if the firm i owns a patent in year t, and 0 otherwise. The OLS regressions do not include any fixed effects, whereas the within regressions included firm fixed effects η_i. Since the sample does not include any other type of firm, the regression coefficients in this table can be interpreted as changes in firm characteristics after patenting (controlling for industry-year effects in the OLS case, and additionally controlling for firm fixed effects in the within regressions).

An inspection of the results in Table 7 suggests that the starkest change (in terms of absolute magnitude) is in the size variables. In all regressions, output, value added, capital stock and employment are higher by about 0.45 to 0.60 log points after the firm switches to a patent-owning status. This translates to an increase in size by a factor of about 1.7 (OLS regressions) to about 1.8 (within regressions) over the pre-patent size, or about 0.2 to 0.3 standard deviations in the overall size distribution. This increase in size is very significant economically, and strongly suggests that a change in patenting status is associated with important changes within the firm. Interestingly, these changes appear to occur mostly around the time of the switch (note that the coefficients in columns 2 and 4 are only marginally less than those in columns 1 and 3).

Turning to differences in skill and capital intensity, the changes are, by and large, consistent with the results in Table 6. Subsequent to patenting, firms generally seem to become more skill-intensive and capital-intensive. However, the magnitude of these differences is much smaller than differences in the key size variables. The white- to blue-collar worker ratio as well as the white-collar share of the wage bill increase by only about 2%. Capital stock per worker shows an increase of about 7 to 7.5%.

Finally, most measures of productivity show no or small changes subsequent to patenting, and the sign of the changes vary depending on the measure used. The OLS regressions show no increase in labor productivity. The OLS TFP results show a small increase in the Solow residual TFP measure; however, the ACF based productivity exhibits a small decrease. With firm fixed effects, the results are only a little bit stronger for labor productivity (a 2% increase around the time of the switch). The TFP results are again mixed.

In summary, patenting appears to be associated with a large increase in size, small increases in skill and capital intensity, and almost no to little improvement in productivity. Combined with the large cross-sectional differences (in most variables) found in Table 6, these results strongly suggest that firms that eventually own patents generally enter with higher productivity, and greater skill and capital intensity. The changes in firm characteristics following the switch is examined in more detail in the next section.

5.2 Event-study (matched cohort) analysis of first-time patenting

While the above analyses show a significant increase in size (and small increases in some of the other variables) subsequent to patenting, it does not provide detailed evidence on the timing of this increase relative to the patenting event itself. Furthermore, the analyses thus far have not been conditioned on any variables, except the industry-year fixed effects. For instance, it is possible that skill intensity (or capital intensity) of firms increases with firm age. Firm size results could also be affected by age, as conditional on survival, firms generally grow in size. Hence, simple "before patenting-after patenting" regressions like those in Table 7 could show an increase in firm size after patenting, solely because firm age has not been factored in the analyses. Another potential source of bias is cohort effects which arise if firms in some size cohorts grew faster than others and the switchers mostly belonged to those cohorts. This section strengthens the conclusions reached in the previous section by addressing some of these gaps, and presents interesting evidence on the timing of changes relative to the patenting event.

To get an idea of the timing of the effects discussed in the previous section, a variable INDEX is defined for each firm, as the difference between the current year, and the first year it applies for a patent. (Note that even though the firm data cover only the five Censuses, the year of patenting is known accurately, as long as the first year of patenting was after 1963.) Hence, INDEX takes negative integer values for the years before the firm first patents, 0 in the first year of patenting, and positive integer values subsequent to its first patent. Since the number of observations for extreme values of INDEX were small, the variable was censored to -15 on the lower end, and to +15 on the upper end, thus allowing INDEX to take any integer value from -15 to +15. This variable INDEX was used to create 31 dummy variables, one for each potential value of the variable, which were subsequently used in OLS regressions of the following form:

$$\Phi_{it} = \sum_{k=-15}^{15} \beta_k D_{kt} + \epsilon_{it} \tag{1}$$

where Φ_{it} is the variable of interest. The sample in these regressions is restricted to firms that switch from non-patentee to patentee status, and D_{-15} is set as the intercept, so that the remaining 30 β_k coefficients offer an idea of how much the variable of interest changes every year relative to its value in the (censored) 15th year before patenting. Hence, a β_0 coefficient of 0.5 on an (log) output regression would imply that in the year of patenting, the average output of a firm is 0.5 log points more than the average output 15 years before. Accordingly, trends in β coefficients with log output or employment as a dependent variable provide a picture of the average firm's growth over the (censored) 30-year period covering

15 years before, and 15 years after patenting.[10]

In order to condition out the effects of industry, size, age and year effects, we use a matched cohort as a control group. Specifically, for every switching firm, a matched control cohort group was defined as the set of firms that did not patent (at least till 1997) which belong to the same industry, size decile and age, as the switching firm in the Census year before patenting. For example, suppose a firm (in SIC 3571) patented for the first time in 1985. The last Census before 1985 was 1982, and suppose that in this Census, the switching firm was 10 years old, and was in size decile 4. Then, the control cohort for this firm would be the set of 10-year old (in 1982) firms that did not patent at least 1997, and were in industry 3571, and in size decile 4 in 1982. The underlying idea here is to identify firms that are very similar to the switching firm before the patenting firm, but did not choose to patent, and to compare the switching firm to these firms.

Using this control cohort, a "differenced variable" was defined as the difference between the variable of interest, and the mean value of that variable for the control cohort. For instance, in the example above, "differenced output" for the switching firm in 1987 would be the difference between its output in 1987, and the mean output in 1987 of survivors in its control cohort (who would now be 15 years old, just as the switching firm would be). The "differenced variable" was used as the dependent variable in regressions identical to Equation 1.

Coefficients from these regressions are plotted in Figure 3. The first set of graphs relate to size variables - output, value added, capital stock and employment. Consistent with the findings in the previous section, the average size before patenting is lower by about 0.6 log points than the average size subsequent to patenting. More interesting, however, is the jump in size *in the year* of first patenting. While there is almost no or little observable growth in the years before patenting, there is a sharp jump of about 0.5 log points in the year of patenting, accounting for a substantial part of the size increase associated with patenting. Subsequent to patenting, firm size appears to grow but at a much slower rate than in the year of patenting. Output shows only a mild increase in the years after patenting (about 1% per year); employment, capital stock, and value added do only marginally better, increasing by about an average of 2 to 3% per year. In contrast, average firm size in the year of patenting jumped by at least 50%, both economically and statistically significant. This implies that

[10] For a similar event study approach, see for example Dinardo and Lee's (2004) study of the impact of unionization on firm outcomes.

even when compared to a set of firms that are very similar to the switchers, a change in patent-ownership status is associated with a large increase in firm size.

The next figure presents changes in skill intensity (as measured by average wages, non-production to production worker ratio, and the wage share of non-production workers). The two skill-intensity variables appear to have a trend broadly more similar to the size variables, exhibiting a small jump in the year of patenting. However, unlike the size variables, this increase appears to continue for a few years after patenting. Note that the magnitude of changes in the skill intensity variables are substantially smaller than the changes in firm size.

The third figure examines changes in capital intensity. Curiously, there is an increasing pattern in capital intensity about 5 to 6 years before patenting, but no change in the period around patenting. This suggests that patenting firms have relatively higher capital intensity some years before patenting relative to non-patenting firms of similar size and age. One potential explanation is that the research activity leading to patentable innovations require relatively higher capital investment per worker starting a few years before commercialization and paptenting of the innovation.

The last graph plots various productivity measures. The trends are broadly similar to the results for capital intensity. There appears to be an increase in labor productivity (output per employee) and the TFP variables a few (5 to 6) years prior to patenting. There appears to be no significant increase in any of the productivity variables around the time of patenting. Also, interestingly, there seems to a small downward trend in the TFP variables after patenting, so that the average TFP levels over the ten year periods before and after the patenting event appear so be similar.

We conclude that the main robust result is that patenting is contemporaneously associated with a large increase in firm size. Most of the "before-after" difference in size from Table 7 appears to be associated with changes in the patenting year. The results here suggest a small increase in skill intensity around the patenting year. While there is some increase in capital intensity and productivity prior to patenting, there is no jump associated with the patenting event itself.

6 Patenting and firm-scope

One potential explanation for the increase in size of the firm around the time of patenting is that patents increase the scope (i.e. number of products) sold by the firm (e.g., as in the model in Klette and Kortum, 2004). In this section, we examine this explanation by defining the scope of the firm as the number of distinct 7 digit product codes reported by it. While this measure is somewhat crude (as there could be significant heterogeneity within 7 digit product codes), it uses the most detailed information on product categories available in the Census data.[11]

In Panel 1 of Table 8, we present the mean difference in the log number of products between patentee and non-patentee firms. As seen here, patentee firms report significantly greater number of products compared to non patentees. Controlling for industry-year effects (Column 3), patentees report about twice ($e^{0.697}$) as many product categories as non-patentees. In Panel 2, we look at the sample of "switchers" (as in Table 7). We find that firms that switch from patentee to non-patentee status do indeed increase the number of product offerings significantly. The results from the fixed effects (within) regressions suggest a 6.4 percent increase in the number of products offered around the time of the switch.

These findings are confirmed by the event-study analysis in Figure 5. The figure suggests a 8 per cent jump in number of products around in the year of patenting.

We conclude that the there is indeed a sizeable increase in the scope of the firm following patenting, and the coincidence of this increase with the increase in size suggests that in our data of switching firms, the patenting event is coincident with the successful launch of new products.[12]

7 Robustness checks

In this section, we discuss a number of robustness checks that we undertook to confirm our findings.

[11] Also, data on 7 digit product codes is missing for a number of firms. We found that the missing data is concentrated within a few industries, and had a similar pattern across patentees and non-patentees. Hence, we do not expect the missing data to bias our results in any systematic way.

[12] In unreported results, we looked at number of new products (defined as the number of previously unreported 7 digit product codes) and find a similar pattern as we found for total number of products here. This confirms that the increase in total products indeed reflects new products and not a re-launch of products that the firm may have previously dropped from its portfolio.

7.1 Measurement error from acquisitions

Since all our analysis is at the firm (rather than establishment) level, it is possible that the "switchers" (analyzed in Sections 5 and 5.2) consist largely of firms that acquire another patent-owning firm rather than patent on their own. If this were true, it could largely explain the sudden jump in firm size and firm scope, contemporaneous with the patenting event.

As a test of robustness, firms that appeared to acquire a patent-owning firm were excluded (recall that not all patent assignees were matched, and hence, this may not completely eliminate all such acquisitions), and the within regressions in Table 7 repeated on this sample. These results are presented in column 1 of Table 9, and are very similar to those in Table 7. As a stricter test, the sample was restricted to firms that consisted only of one establishment. These results are presented in column 2 of Table 9, and support the previous findings.

In unreported tests, firms that appeared to enter with more than 5 patents were excluded (since a large number of patents could signal an unobserved acquisition), and the within regressions re-estimated. The results were similar.

7.2 Elasticity of firm characteristics to patent stock

Our results suggest a significant increase in firm size and scope when firms switch status from non-patentee to patentee. Given this result, we would expect that increases in the stock of patents would have similar effects, as we expect additional patents to be associated with further increases in the number of products produced by the firm.[13] We examine this by regressing the dependent variable on log of depreciated patent stock.[14]

The results are summarized in Table 10. In column 1, the sample includes all patentees (with dependent variables demeaned for industry-year effects). In column 2, to make the sample comparable to the one used in Table 7, we include only "switchers", i.e. firms that enter our panel as non-patentees and then switch status to patentees. Note that since the pre-patenting log patent stock is undefined, and since all regressions include firm fixed effects, these regressions examine the relationship between changes of firm characteristics within patentee firms (with at least one patent) following increases in the log patent stock.

[13]The measured impact of the first patenting event may be higher if there are significantly high fixed cost associated with patenting for the first time (for example in understanding the patenting process), which are absent for subsequent patenting.

[14]We use a depreciation rate of 15% following Hall, Jaffe and Trajtenberg, 2005.

As expected, there is a significant and sizeable elasticity of size with respect to patent stock. A 10 percent increase in log patent stock is associated with about 1.8 percent increase in size (whether measured as output, value added, capital stock or employment). The magnitudes of the effects are slightly smaller (13.5 to 15 percent) in the sample of switchers alone (column 2). The effect on skill intensity is not significant (except the white- to blue-collar worker ratio which shows marginally significant increase with patent stock in the sample of all patentees). Similarly, capital intensity is marginally significant in the full sample, but insignificant in the sample of switchers only. None of the productivity measures show statistically significant increases with increase in patent stocks.

Finally, the number of products produced by the firm is significantly related to the patent stock. A 10% increase in patent stock is associated with a 1.4% (1% in the sample of switchers only) increase in the number of products offered by the firm.

We conclude that the analysis of the responsiveness of firm characteristics to patent stock confirms the main findings in our earlier analysis that: (i) patenting is associated with an increase in firm size; (ii) this increase is size appears to be driven by an increase in firm scope; and (iii) there appears to be no significant increase in productivity associated with patenting. The results here suggest that the findings in Section **??** of an increase in skill intensity post-patenting may be less significant for subsequent patenting within the firm. Similarly the effect on capital intensity of increases in patent stock within patentees appear to be small.

7.3 Link to patent quality

If the observed increase in size is indeed linked to patenting, and not due to other unrelated causes, it is reasonable to expect a link between the size of the increase, and the "quality" of the patents. In order to test this, the "differenced output" variable (demeaned of matched cohort effects), was regressed on the patent-ownership dummy, a quality measure, and the interaction of these two variables. Three measures of quality were used: (i) the number of patents in the first year of patenting; (ii) the total number of forward citations to the patents in the first year of patenting; and (iii) the number of citations per patent for patents in the first year of patenting. These results are presented in Table 11. For all three measures of innovation quality, the interaction of the patent-ownership dummy and the quality variable is significantly positive. These results support the conclusion that the observed increase is likely due to patenting rather than other unrelated causes.[15]

7.4 Other checks

Another potential problem is that the results are being driven by a few industries, or a few years. In order to rule this out, the within regressions were repeated by excluding one SIC2 industry at a time: The results were very similar to the baseline specifications. Repeating the tests, excluding one year at a time, produced similar results too.

In addition to these statistical tests, a visual inspection on a small sample of single establishment "switching" firms was performed to check if these firms do show an increase in employment in the year of patenting. A majority of these firms did indeed exhibit an increase, with a small minority displaying substantial increases. Though not conclusive, this test is also consistent with the results. Finally, in order to rule out other potential matching problems, the tests in Table 7 were repeated with only those assignees matched with the highest reliability code (see discussion in Appendix 1). The results remained similar.

8 Discussion of findings

The main robust finding from our analyses is that (i) patenting is associated with a significant increase in size of the firm; and (ii) this increase in size is coincident with an increase in the number of products produced by the firm. Other results are mixed. Our event study analysis suggests a small increase in capital intensity and productivity a few years prior to patenting, and a modest jump in skill intensity coincident with patenting, but these results disappear when we look at changes in firm characteristics with respect to patent stock (Table 10). Further, results on total factor productivity for first time patentees are mixed, with negative changes in TFP when we use the Ackerberg-Caves-Frazer measure (Tables 7 and 9).

Griliches (1990) proposed that, ideally, we would like patent statistics "to measure and better understand the economic processes that lead to the reduction in the cost of producing existing products and the development of new products and services". Our results suggest that, at the firm level, patents are more reflective of the development of new products, rather than reductions in the cost of existing products (which should be captured by our TFP measure). Our finding of an increase in scope but little or no increase in total factor

[15]Consistent with previous results, repeating these quality tests with productivity did not produce any significant coefficients on the interactions.

productivity is consistent with the findings in Levin et al (1987). They find that patents are viewed as effective mechanisms for appropriability much more for product rather than process patents. They find that secrecy was much more effective for process innovations, which suggests that many TFP improving process innovations may not be patented at all.

The coincidence of the increase in size and scope of the firm with patenting suggests that most firms file for patents simultaneously with the launch of a new product. Thus, firms appear to depend on secrecy initially and resort to patenting only around the time of the new product launch. This is also consistent with the finding in Levin et al (1987) and Cohen et al (2000) that in many industries, patents may not be very effective in protecting an innovation, and that patents may themselves be a source of spillovers to competitors, as rivals learn about the innovation from the patent filing and "invent around" patents.

Our results suggest that the recent class of models that explain firm heterogeneity based on differences in firm scope (such as Klette and Kortum, 2004; Bernard, Redding and Schott, 2006b; Nocke and Yeaple, 2006) may be more suitable to understanding patenting at the firm level than older models that focused on productivity as the source of firm heterogeneity (e.g. Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes, 1995; Melitz, 2003). Our results are broadly consistent with the model of patenting presented in Klette and Kortum (2004). In this model, a firm is defined by the portfolio of goods it produces, and it innovates by extending its product line. Innovations are not directly linked to productivity; hence, assuming patents represent successful innovations, the scope (but not necessarily the productivity) of a firm increases coincident with patenting. In this model, the size of the innovative steps is positively related to measured productivity and to R&D intensity. Since arrival rate of innovations (and hence patenting) is a positive function of R&D intensity, ex-ante differences in productivity between patenting and non-patenting firms would be consistent with the model.[16]

Interestingly, the results on changes in firm size and productivity that we document for patenting are similar to the empirical results on firm size, productivity and exporting documented by Bernard and Jensen (1999) and Clerides, Lach and Tybout (1998) in the trade literature. These papers document that overall exporters are more productive than non-exporters, but find little evidence of an increase in total factor productivity subsequent

[16] Additional assumptions that not all innovations are patented and that patenting involves a fixed cost but provides some benefits in the form of extra protection for markups may be needed to explain ex-ante size and productivity differences between patentees and non-patentees and to explain why all firms do not patent.

to a change in status from non-exporting to exporting. To explain this finding, Melitz (2003) and Bernard, Jensen, Eaton and Kortum (2003) have proposed heterogeneous firm equilibrium models where cross-sectional differences in productivity between exporters and non-exporters arise as a result of self-selective entry by productive firms into foreign markets. A key part of the explanation is that there is a fixed cost to access foreign markets, which makes foreign entry affordable only for the more productive firms. Modifications of these models, with a fixed cost of increasing number of products (and/or for filing a patent) could explain the positive overall correlation between productivity and patenting, as well as the lack of increase in TFP post-patenting. To extend the analogy, new products could be considered similar to new markets that a firm could access.[17]

One limitation of our work is that we lack specific information on potential investments by firms prior to patenting.[18] Without good information on the costs, we cannot assess if the increase in size and scope that we document reflect increase in welfare or even increase in value of the firm. In addition to data on costs, evaluating potential welfare gains from introduction of new products requires data to estimate cross-price elasticity of demand between the new products and old products, which is also unavailable.

9 Conclusions

In this study, we contribute to the rich literature on patenting in three ways.

First, we significantly extend the available patent-firm data link, by matching corporate assignees in the NBER patent data set to firms in the US Census microdata. Our linked dataset improves on the existing firm-patent data link in a number of ways. One, it significantly expands the linkage between patent and firm data (from about 4100 in the existing Compustat link to about 48,000 unique assignees). Two, the linkage to census datasets provides detailed input and output data for the entire universe of US (employer) firms, including unlisted firms. In particular, firm level data on wage and employment available here are either unavailable or of poor quality in Compustat or other financial statement data. Three, our annual link to the SSEL allows tracking ownership changes over time.[19] Four,

[17] An important difference from exporting may be the greater uncertainty in new product development as well as the crucial role for R&D investments.

[18] It is likely that for most firms a significant part of labor and capital resources committed to innovation are captured in the operating expenses reported in the census data. However, we do not have a breakup of the costs incurred by the firms specifically in activities that led to the innovation that was patented.

[19] Researchers at the US Census has undertaken enormous efforts in improving the quality of ownership links over time, first through the Longitudinal Research Database, and then through the Longitudinal Business Database (Jarmin and

the availability of data on both patentees and non-patentees allows for the use of matched or random control groups, which may be unavailable in analysis that uses only the NBER data. We believe our linked dataset will help future researchers to explore a variety of questions relating to patenting, innovation and growth, and hence contribute to these active and important areas of empirical research.

Second, we use our linked data to present a number of new descriptive statistics on the prevalence of patenting in the US manufacturing sector. These statistics provide an independent confirmation of the skewed nature of patenting; in the manufacturing sector as a whole and within sub-sectors, a small minority of firms account for all the patenting activity, and these firms account for a disproportionately large share of economic activity. A particularly interesting empirical fact we document, worthy of further study, is that the share of patenting firms in aggregate employment decreased significantly between 1977 and 1977, even as the share of aggregate output and value added increased considerably in the same period.

Third, we use the linked data to address a fundamental question posed in the survey by (Griliches 1990) - what aspects of economic activity do patent statistics actually capture? - by analyzing what happens within firms when they patent. We find that a change in patenting status is contemporaneously associated with a large increase in firm size, and little or no change in total factor productivity (and other firm characteristics). We also find that there is a large contemporaneous increase in the number of products offered by the firm. These findings suggest that patents are associated mainly with new product launches that increases firm size, consistent with multi-product models like Klette and Kortum (2004) but in contrast to productivity based models such as Melitz (2003) or Hopenhayn (1992). To our knowledge, ours is the most comprehensive examination to date of the differences between patenting and non-patenting firms, and the first to use a large, representative sample of first time patentees to address the question of what happens when firms patent. We hope our findings (and those documented by future research using our link) will inform the development of theoretical models of innovation and patenting.

The timing of the increase in size (coincident with the filing of the patent) suggests that filing of patents is indeed associated with real changes at the firm level. An interest-

Miranda, 2002). The existing link between the patent dataset and Compustat is based on available data for 1989, making it difficult to track certain types of patentees. For instance, first-time patents even by listed (Compustat) firms after 1989 would not be reflected in the currently available patent data-firm data link. Also, acquisition of an unlisted patent-owning firm by a listed non-patentee firm would not be reflected in the current link.

ing extension would be to examine closely the trends over time in the impact of patents (especially over the last decade), especially in the context of regulatory changes (e.g. the establishment of the court of appeals in 1982, the Hatch-Waxman Act of 1984), the large increase in patent filing in recent years (Kortum and Lerner, 1998), and recent concerns regarding patent thickets (Shapiro, 2001) and patent trolling (Barker, 2005). Examining cross industry variations in the impact of patenting on size and scope could be another interesting extension, as the importance (and prevalence) of patenting varies considerably across industries.

References

[1] Ackerberg, D., Caves, K. & Frazer, G. 2005. Structural estimation of production functions: An application to the timing of input choices. Working Paper, University of California, Los Angeles, August 2005.

[2] Austin, D.H. 1993. An event-study approach to measuring innovative output: The case of biotechnology. *American Economic Review*, 83(2), 253-258.

[3] Balasubramanian, Natarajan, 2007. Essays on learning and innovation. Unpublished Dissertation, University of California, Los Angeles.

[4] Bahk, B.H. & Gort M. 1993. Decomposing learning-by-doing in new plants, *The Journal of Political Economy*, 101: 561-583.

[5] Bartelsman, E.J., Becker, R.A. & Gray, W.B. 2000. NBER-CES Manufacturing Industry Database. June 2000.

[6] Barker, D.G. 2005. Troll or no troll? Policing patent usage with an open post-grant review. *Duke Law and Technology Review*, No 9.

[7] Bernard A.B & Jensen J.B. 1999. Exceptional exporter performance: cause, effect, or both? *Journal of International Economics*, 47 (1): 1-25.

[8] Bernard, A. B., Eaton, J., Jensen, J. B. and Kortum, S., 2003. Plants and productivity in international trade. *American Economic Review*, 93(4), 1268-1290.

[9] Bernard A.B, Redding S., & Schott, S. 2006a. Multiproduct firms and product switching. Working Paper, Tuck School of Business at Dartmouth, May 2006.

[10] Bernard A.B, Redding S., & Schott, S. 2006b. Multiproduct firms and trade liberalization. Working Paper, Tuck School of Business at Dartmouth, December 2006.

[11] Bloom N. & Van Reenen J. 2002. Patents, Real Options and Firm Performance. *Economic Journal*, 112, pp. C97-C116.

[12] Branstetter L.G., Fisman R, & Foley C.F. 2006. Do stronger intellectual property rights increase international technology transfer? Empirical evidence from US firm-level panel data. *Quarterly Journal of Economics*, 121(1): 321-349.

[13] Chiang, H., 2004. Learning by Doing, Worker Turnover, and Productivity Dynamics. University of Maryland, mimeo.

[14] Clerides, S. K., Lach, S. and Tybout, J. R., 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics*, 113(3), 903-947.

[15] Cohen, W. M. and Klepper, S., 1992. The Anatomy of Industry Research-and-Development Intensity Distributions. *American Economic Review*, 82(4), 773-779.

[16] Cohen, W. M. and Klepper, S., 1996. A reprise of size and R&D. *Economic Journal*, 106(437), 925-951.

[17] Cohen, W. M., Nelson R. R. and Walsh, J.P., 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). NBER Working Paper No. 7552.

[18] Davis, S., Haltiwanger, J. and Schuh, S., 1996. *Job Creation and Destruction*. Cambridge: MIT Press.

[19] Dinardo, J. and Lee, David S., 2004. Economic impacts of unionization on private sector employers: 1984-2001. NBER working paper 10598.

[20] Ericson, R., and Pakes, A., 1995. Markov-Perfect Industry Dynamics - A Framework for Empirical Work. *Review of Economic Studies* 62, 53-82.

[21] Griliches, Z. 1981. Market value, R&D and patents, *Economics Letters*, 7, 183-87.

[22] Griliches, Z., 1990. Patent Statistics as Economic Indicators - A Survey. *Journal of Economic Literature* 28, 1661-1707.

[23] Griliches, Z. & Mairesse, J. 1995. Production functions: the search for identification. Working paper no. 5067, National Bureau of Economic Research.

[24] Hall, B.H, Jaffe, A.B. & Trajtenberg, M. 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. NBER working paper 8498.

[25] Hall BH, Jaffe A, Trajtenberg M. 2005. Market value and patent citations. *Rand Journal of Economics*, 36 (1): 16-38.

[26] Hall, B.H. & Ziedonis R.H. 2001. The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995. *Rand Journal Of Economics* 32(1): 101-128

[27] Hopenhayn, H. 1992. Entry, exit and firm dynamics in the long run. *Econometrica*, 60(5): 1127-1150.

[28] Jarmin, R. S., and Miranda, J. 2002. The Longitudinal Business Database. Working Papers 02-17, Center for Economic Studies, U.S. Census Bureau.

[29] Jaffe, A. B., Trajtenberg, M. and Henderson, R. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, 108(3), 577-598.

[30] Jovanovic, B., 1982. Selection and the Evolution of Industry. *Econometrica* 50, 649-670.

[31] Kerr, W. & Fu, S. 2006. The Industry R&D Survey: Patent Database Link Project. CES Working Paper CES-WP-06-28.

[32] Klette, T. J. and Kortum, S., 2004. Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5), 986-1018.

[33] Kortum, S. and Lerner, J., 1999. What is behind the recent surge in patenting? *Research Policy*, 28(1), 1-22.

[34] Kortum, S. and Lerner, J., 2000. Assessing the contribution of venture capital to innovation. *Rand Journal of Economics*, 31(4), 674-692.

[35] Levin, R. C., Klevorick, A. K., Nelson, R. R. and Winter, S. G., 1987. Appropriating the returns from industrial-research and development. Brookings Papers on Economic Activity (3), 783-831.

[36] Levinsohn, J. & Petrin, A., 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, Blackwell Publishing, vol. 70(2), pages 317-341, 04.

[37] Mann R.J. & Sager T.W. 2007. Patents, venture capital, and software start-ups. *Research Policy*, 36(2): 193-208.

[38] Melitz, M. J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695-1725.

[39] Miller, D. J., 2006. Technological diversity, related diversification, and firm performance. *Strategic Management Journal*, 27(7), 601-619.

[40] Nocke, V. and Yeaple, S. 2006. Globalization and Endogenous Firm Scope. NBER Working Paper Number 12322, National Bureau of Economic Research

[41] Olley,G.S. & Pakes, A. 1996. The dynamics of productivity in the telecommunication equipment industry. *Econometrica*, 64(6): 1263-1297

[42] Pakes, A. 1985. On patents, R&D and the stock market rate of return. *Journal of Political Economy*, 93(2): 390-409.

[43] Sakakibara, M. & Branstetter, L. 2001. Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms. *Rand Journal of Economics*, 32(1): 77-100.

[44] Sampath, B.N., & Ziedonis, A. 2004. Patent Citations and the Economic Value of Patents: A Preliminary Assessment. In *Handbook of Quantitative Science and Technology Research*, Henk Moed, and Wolfgang Glanzel, Ulrich Schmoch, eds., (Kluwer Academic Publishers: Boston, MA).

[45] Scherer, F. M., 1965. Firm Size, Market-Structure, Opportunity, and the Output of Patented Inventions. *American Economic Review* 55, 1097-1125.

[46] Scherer, F. M., 1983. The Propensity to Patent. *International Journal of Industrial Organization* 1, 226-37.

[47] Shapiro, C. 2001. Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting. In *Innovation Policy and the Economy*, A. B. Jaffe, J. Lerner and S. Stern eds., vol.1, MIT Press, pp. 119-150.

TABLE 1: MATCHED DATASET – OVERALL COVERAGE

EVER-MATCHED	Patents	Citation-Wtd Patents	Assignees
1963-1999: All Assignees	0.403 (2.92 m)	0.479 (14.0 m)	0.329 (175,115)
1963-1999: US Non-Individual, Non-Government Assignees Excluding Universities	0.899 (1.07 m)	0.908 (6.20 m)	0.599 (95,157)
1975-1997: US Non-Individual, Non-Government Assignees Excluding Universities	0.900 (0.78 m)	0.910 (4.54 m)	0.637 (76,380)

CONTEMPORANEOUS MATCH	Patents	Citation-Wtd Patents	Assignee-years
1975-1997: US Non-Individual, Non-Government Assignees Excluding Universities	0.806 (0.78m)	0.797 (4.54 m)	0.640 (189,835)

Notes: (i) Each cell presents two numbers. The first is the fraction of patents or of assignees covered in the match. The second, in parentheses, provides the total number of patents or assignees in the relevant population. (ii) The match coverage above includes patent assignees not in the manufacturing sector. (iii) "Ever-Matched" is defined as a Patent Assignee-SSEL Firm match, regardless of the year of patenting. A "Contemporaneous Match" is defined as a Patent Assignee-SSEL Firm *in the (application) year of patenting*. Hence, in any given year, a contemporaneous match exists only if the assignee applied for a patent during that year, and can be found in the SSEL. (iv) "m" represents million.

FIGURE 1: NBER-CENSUS LINKED DATASET -- YEAR BY YEAR COVERAGE

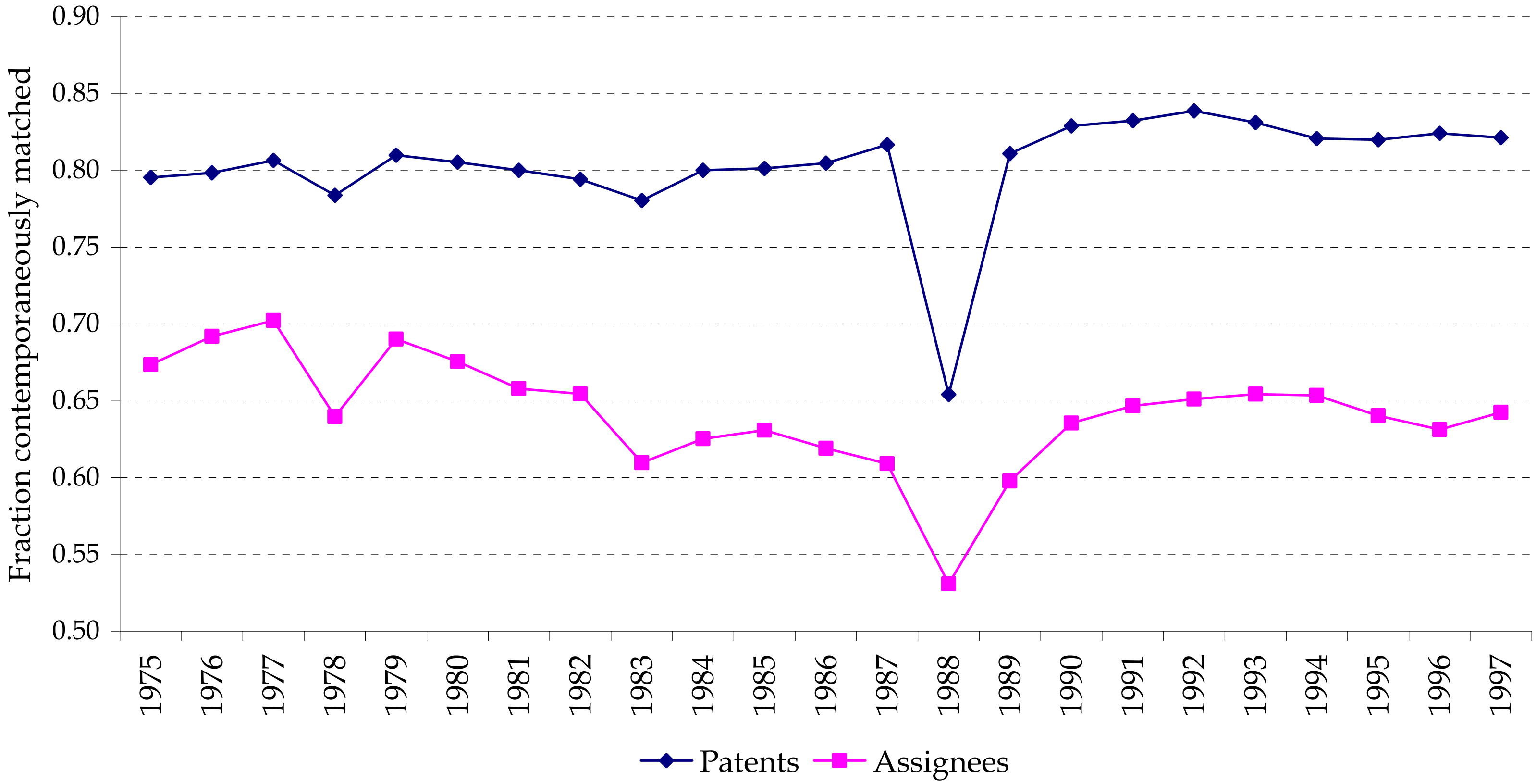


TABLE 2: MATCHED DATASET –COVERAGE BY TECHNOLOGICAL CATEGORY

	Ever-Matched			Contemporaneous Match		
	Patents	**Citation-Wtd Patents**	**Assignees**	**Patents**	**Citation-Wtd Patents**	**Assignees**
Chemicals (1)	0.919	0.929	0.699	0.831	0.828	0.688
Computers & Communication (2)	0.925	0.923	0.700	0.833	0.791	0.676
Drugs & Medical (3)	0.856	0.871	0.635	0.750	0.726	0.615
Electrical & Electronic (4)	0.926	0.931	0.719	0.845	0.837	0.714
Mechanical (5)	0.894	0.906	0.702	0.799	0.799	0.694
Others (6)	0.862	0.879	0.674	0.756	0.763	0.662

Notes: The sample covers only patents applied during 1975-1997, and belonging to US non-individual, non-government assignees that are not universities

TABLE 3: COMPARISON OF MATCHED AND NON-MATCHED ASSIGNEES

	Ever-Matched		Contemporaneous Match	
	Matched	**Non-Matched**	**Matched**	**Non-Matched**
Number of Patents	14.42	2.81	1.73	1.27
Number of Citation-Wtd Patents	84.89	14.69	11.71	8.02
Number of Citations per Patent	5.21	5.10	6.22	6.20
Number of Claims per Patent	13.45	13.50	13.60	13.94

Notes: (i) The relevant population is US non-individual non-government assignees, excluding universities. (ii) The match coverage includes patent assignees not in the manufacturing sector. (iii) For "Ever-Matched" analyses, data across all years are used. For "Contemporaneous Match" analyses, data only in the year of application is used. Hence, the number of patents and citation-weighted patents is higher in the "Ever-matched" analyses.

TABLE 4: ECONOMIC ACTIVITY IN MANUFACTURING - SHARE OF PATENTEES (BY SIC2 INDUSTRY)

		Number of firms	Value Added	Capital Stock	Employment
20	Food and kindred products	1.9%	60.5%	61.7%	48.1%
22	Textile mill products	3.7%	51.5%	56.4%	48.3%
23	Apparel and other textile products	0.9%	26.5%	34.0%	19.5%
24	Lumber and wood products	0.5%	29.4%	37.6%	20.2%
25	Furniture and fixtures	2.7%	42.6%	48.3%	34.3%
26	Paper and allied products	7.1%	77.2%	82.2%	70.6%
27	Printing and publishing	0.5%	33.4%	35.3%	24.4%
28	Chemicals and allied products	9.2%	86.1%	88.1%	79.8%
29	Petroleum and coal products	7.0%	72.4%	84.2%	78.7%
30	Rubber and miscellaneous plastics	7.6%	59.8%	62.9%	51.3%
31	Leather and leather products	3.4%	40.0%	45.7%	33.3%
32	Stone, clay, glass, and concrete products	2.2%	53.8%	56.4%	44.4%
33	Primary metal industries	6.6%	74.2%	80.3%	67.8%
34	Fabricated metal products	5.2%	48.7%	51.5%	41.3%
35	Industrial machinery and equipment	6.8%	72.5%	72.9%	60.5%
36	Electrical and electronic equipment	12.0%	77.4%	77.8%	67.5%
37	Transportation equipment	5.8%	90.3%	90.3%	85.2%
38	Instruments and related products	17.4%	86.8%	89.2%	81.3%
39	Miscellaneous manufacturing industries	3.8%	43.2%	51.2%	35.5%
	ALL INDUSTRIES	**5.5%**	**59.3%**	**63.5%**	**52.2%**

Notes: (i) The numbers represent averages over the period 1977 to 1997. (ii) Due to disclosure restrictions, the statistics for SIC 21 (Tobacco manufactures) have been merged with SIC 20 (Food and kindred products).

FIGURE 2: SHARE OF PATENTING FIRMS IN MANUFACTURING ACTIVITY – BY SIZE CLASS

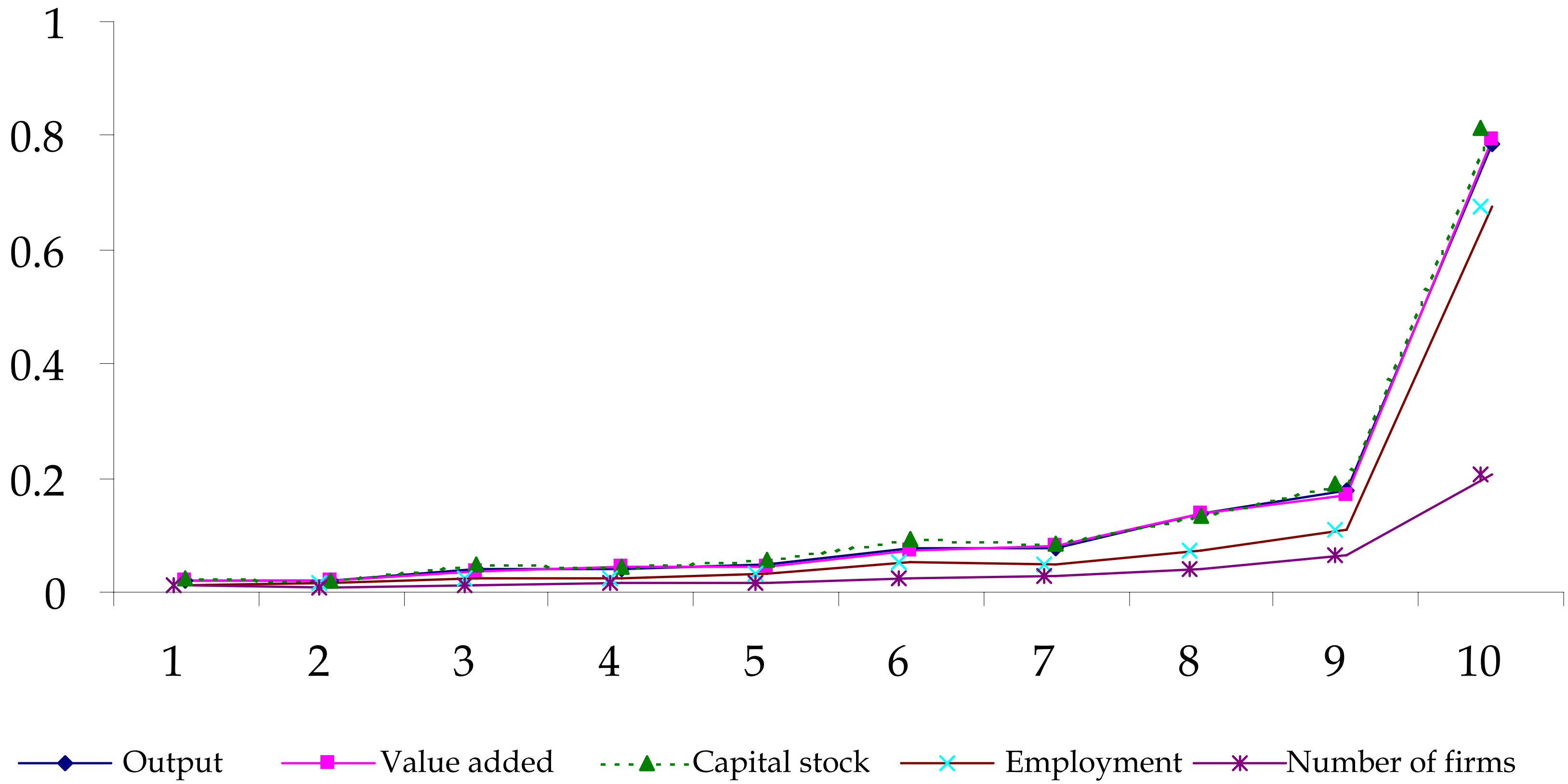


TABLE 5: TEMPORAL CHANGES IN ROLE OF PATENTEES

	1977	1987	1997	Change 1977-87	Change 1977-97
Patent-owning Firms					
Output	1,460,000	1,710,000	2,930,000	250,000	1,470,000
Value added	647,000	847,000	1,710,000	200,000	1,063,000
Capital stock	412,000	676,000	1,020,000	264,000	608,000
Employment	10,500,000	9,850,000	9,310,000	-650,000	-1,190,000
Non-Patent Owning Firms					
Output	680,000	760,000	1,030,000	80,000	350,000
Value added	321,000	383,000	530,000	62,000	209,000
Capital stock	145,000	237,000	390,000	92,000	245,000
Employment	8,000,000	7,850,000	8,190,000	-150,000	190,000
Total					
Output	2,140,000	2,470,000	3,960,000	330,000	1,820,000
Value added	968,000	1,230,000	2,240,000	262,000	1,272,000
Capital stock	557,000	913,000	1,410,000	356,000	853,000
Employment	18,500,000	17,700,000	17,500,000	-800,000	-1,000,000

Notes: (i) Output, value added and capital stock in millions of 1987 dollars. (ii) Patent-owning firms are defined as firms that owned a patent sometime from 1963 to 1997, and non-patent owning firms are firms that never owned a patent.

TABLE 6: DIFFERENCES IN MEAN FIRM CHARACTERISTICS (PATENTEES VS. NON-PATENTEES)

	Patent-Owning Mean	Non-Patent Owning Mean	Dummy Variable Regression
Size			
Output	8.915 (2.19)	6.195 (1.63)	2.292 (0.042)***
Value added	8.329 (2.16)	5.598 (1.63)	2.309 (0.043)***
Capital stock	7.600 (2.36)	4.813 (1.75)	2.339 (0.044)***
Employment	4.378 (1.96)	2.079 (1.43)	2.072 (0.040)***
Skill and capital intensity			
White-collar share of wage bill	0.457 (0.19)	0.346 (0.19)	0.046 (0.002)***
White- to blue-collar worker ratio	0.343 (0.19)	0.233 (0.22)	0.058 (0.002)***
Capital stock per worker	3.678 (1.05)	3.035 (1.10)	0.348 (0.009)***
Productivity			
Labor productivity	4.536 (0.66)	4.116 (0.72)	0.218 (0.006)***
TFP (Solow residual)	2.213 (0.51)	2.056 (0.57)	0.142 (0.004)***
TFP (Ackerberg-Caves-Frazer)	2.129 (0.65)	1.933 (0.64)	0.080 (0.004)***

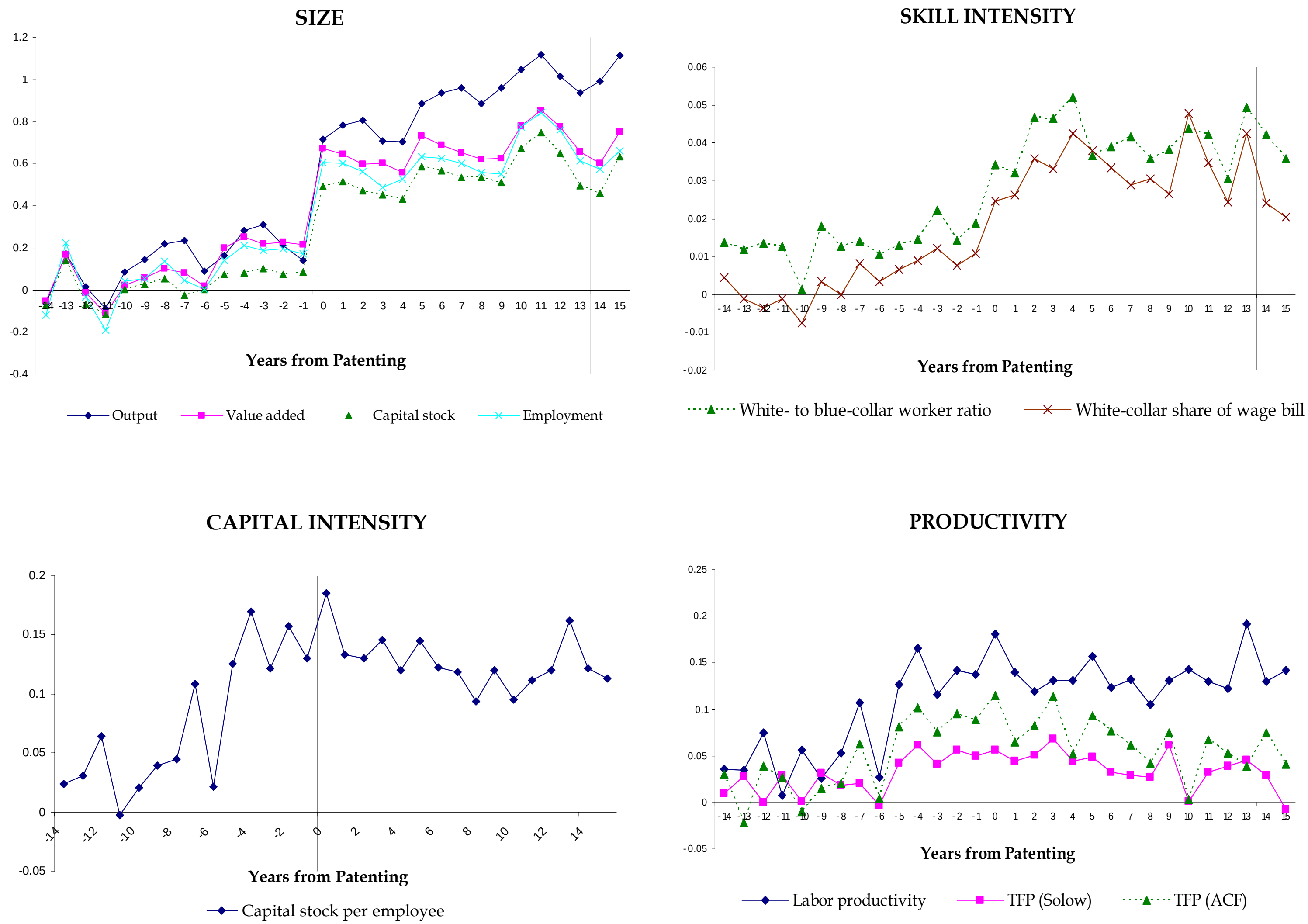
Notes: (i) Patent-ownership is defined as 1 if a firm owns a patent in the current year, 0 otherwise. All dependent variables, except white collar share of wage bill and the white- to blue-collar wage ratio, are logged. (ii) The first two columns of the table present the means for patent-owning and non-patent owning firms, the average taken over 5 census years 1977, 1982, 1987, 1992 and 1997. Standard deviations of the same variable are presented in the parentheses. (iii) The total number of observations used above is around 1.1 to 1.4 million of which about 4.3% belong to patent-owning firms. (iv) Column 3 reports coefficients on a dummy indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. SIC4-year fixed effects are included in all regressions. Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

TABLE 7: DIFFERENCES IN MEAN FIRM CHARACTERISTICS BEFORE AND AFTER PATENTING

	OLS Regressions		Within Regressions	
	Switchers overall	**Around the Switch**	**Switchers overall**	**Around the Switch**
Size				
Output	0.502 (0.032)***	0.460 (0.032)***	0.595 (0.024)***	0.440 (0.027)***
Value added	0.513 (0.032)***	0.466 (0.033)***	0.599 (0.025)***	0.437 (0.030)***
Capital stock	0.543 (0.034)***	0.494 (0.034)***	0.607 (0.025)***	0.453 (0.028)***
Employment	0.507 (0.031)***	0.471 (0.031)***	0.566 (0.024)***	0.420 (0.027)***
Skill and capital intensity				
White-collar share of wage bill	0.019 (0.003)***	0.019 (0.003)***	0.023 (0.003)***	0.027 (0.003)***
White- to blue-collar worker ratio	0.018 (0.003)***	0.020 (0.003)***	0.021 (0.003)***	0.023 (0.004)***
Capital stock per worker	0.068 (0.014)***	0.057 (0.016)***	0.077 (0.013)***	0.075 (0.018)***
Productivity				
Labor productivity	-0.004 (0.009)	-0.011 (0.010)	0.027 (0.008)***	0.020 (0.009)***
TFP (Solow residual)	0.021 (0.006)***	0.016 (0.007)***	0.034 (0.005)***	0.025 (0.007)***
TFP (Ackerberg-Caves-Frazer)	-0.020 (0.008)**	-0.029 (0.009)***	-0.008 (0.008)	-0.020 (0.011)*

Notes: (i) This table reports coefficients on a dummy indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. All dependent variables, except white collar share of wage bill and the white- to blue-collar wage ratio, are logged. (ii) The samples in Column 1 and 3 are restricted to those firms that enter without a patent and switch to being patent-owning firm later in their life ("switchers"). Additionally, the samples in Column 2 and 4 are restricted to the five years before and the five years after the "switch". (iii) The dependent variables are demeaned to control for industry-year effects. Regressions in Columns 2 and 3 include firm fixed effects. (iv) The sample size in these regressions varies from about 15,000 to about 30,000. (v) Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

FIGURE 3: EVENT STUDY (MATCHED COHORTS) ANALYSIS



Notes: i) Points on the graph are coefficients on a dummy indicator variable, which equals 1 if the firm owns a patent, interacted with an index indicating the number of years from patenting. All dependent variables, except white-collar share of wage bill and the white- to blue-collar wage ratio, are logged. (ii) The dependent variable is computed as the value of the variable less the mean of control group, matched on size, age, and industry in the year prior to patenting. (iii) Number of observations in the underlying regressions is about 28,000.

TABLE 8: CHANGE IN FIRM SCOPE (NUMBER OF PRODUCTS)

Panel 1: Differences in mean number of products

	Patent-Owning Mean	Non-Patent Owning Mean	Dummy Variable Regression
Number of products	1.328 (1.20)	0.572 (0.68)	0.697 (0.029)***

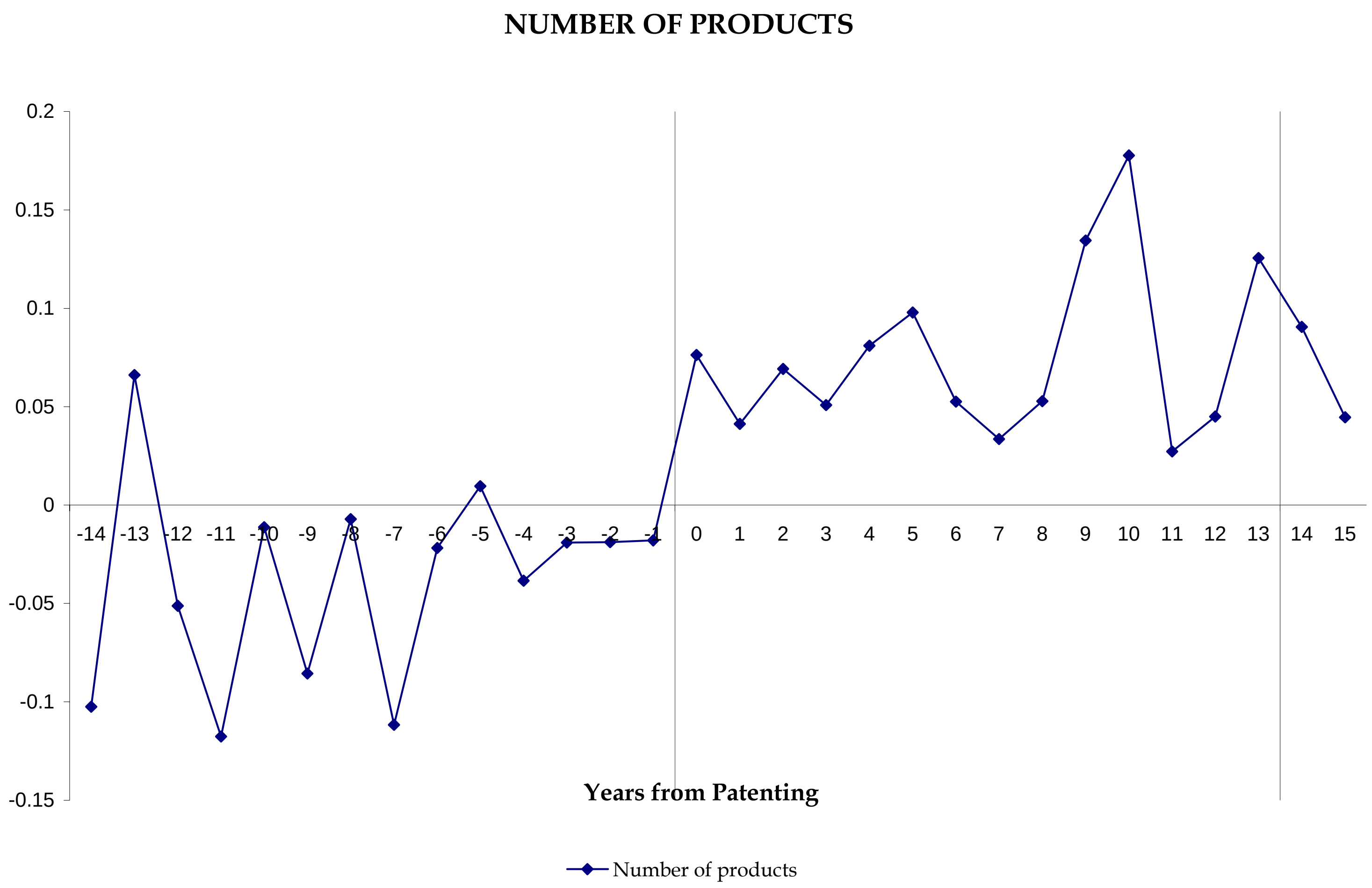
Notes: (i) Patent-ownership is defined as 1 if a firm owns a patent in the current year, 0 otherwise. The dependent variable (number of products) is logged. (ii) The first two columns of the table present the means for patent-owning and non-patent owning firms, the average taken over 5 census years 1977, 1982, 1987, 1992 and 1997. Standard deviations of the same variable are presented in the parentheses. (iii) Column 3 reports coefficients on a dummy indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. SIC4-year fixed effects are included in all regressions. Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.. (iv) Number of products refers to the (log of) number of distinct 7-digit product classes that the firm is present in. (v) The total number of observations used above is around 1.1 to 1.4 million of which about 4.3% belong to patent-owning firms.

Panel 2: Mean number of products before and after patenting

	OLS Regressions		Within Regressions	
	Switchers overall	Around the Switch	Switchers overall	Around the Switch
Number of products	0.209 (0.016)***	0.180 (0.016)***	0.062 (0.010)***	0.064 (0.013)***

Notes: (i) This table reports coefficients on a dummy indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. The dependent variable (number of products) is logged. (ii) The samples in Column 1 and 3 are restricted to those firms that enter without a patent and switch to being patent-owning firm later in their life (“switchers”). Additionally, the samples in Column 2 and 4 are restricted to the five years before and the five years after the “switch”. (iii) The dependent variables are demeaned to control for industry-year effects. Regressions in Columns 2 and 3 include firm fixed effects. (iv) The sample size in these regressions varies from about 15,000 to about 30,000. (v) Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

FIGURE 4: SCOPE OF FIRM – EVENT STUDY (MATCHED COHORTS) ANALYSIS



Notes: i) Points on the graph are coefficients on a dummy indicator variable, which equals 1 if the firm owns a patent, interacted with an index indicating the number of years from patenting. The dependent variable (number of products) is logged. (ii) In event study analysis figure (left), the dependent variables are demeaned to control for industry-year effects. In the matched cohort analysis figure (right), the dependent variable is computed as the value of the variable less the mean of control group, matched on size, age, and industry in the year prior to patenting. (iii) Number of observations in the underlying regressions is about 28,000.

TABLE 9: DIFFERENCES IN MEAN FIRM CHARACTERISTICS
(ROBUSTNESS CHECKS – EXCLUDING ACQUISITIONS AND MULTI-UNIT FIRMS)

	Excluding possible acquisitions	**Excluding multi-unit firms**
Size		
Output	0.587 (0.026)***	0.500 (0.031)***
Value added	0.594 (0.027)***	0.508 (0.033)***
Capital stock	0.591 (0.026)***	0.497 (0.033)***
Employment	0.561 (0.025)***	0.485 (0.031)***
Skill and capital intensity		
White-collar share of wage bill	0.027 (0.003)***	0.033 (0.004)***
White- to blue-collar worker ratio	0.025 (0.003)***	0.030 (0.004)***
Capital stock per worker	0.075 (0.014)***	0.066 (0.021)***
Productivity		
Labor productivity	0.025 (0.009)***	0.015 (0.012)***
TFP (Solow residual)	0.036 (0.006)***	0.034 (0.008)***
TFP (Ackerberg-Caves-Frazer)	-0.006 (0.008)	-0.006 (0.013)
Scope		
Number of products	0.022 (0.011)***	0.062 (0.016)***

Notes: (i) This table reports coefficients on a dummy indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. All dependent variables, except white-collar share of wage bill and the white- to blue-collar wage ratio, are logged. (ii) The samples are restricted to those firms that enter without a patent and switch to being patent-owning firm later in their life ("switchers"). Column 1 excludes firms that appeared to acquire a patent owning firm. Column 2 excludes all multi-unit firms. (iii) The dependent variables are demeaned to control for industry-year effects. All regressions include firm fixed effects. (iv) The sample size in these regressions varies from about 15,000 to about 30,000. (v) Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

TABLE 10: ELASTICITY WITH RESPECT TO PATENT STOCK (WITHIN REGRESSIONS)

	All patentees	Switchers only
Size		
Output	0.180(0.011)**	0.141(0.020)**
Value added	0.179(0.0109)**	0.137(0.019)**
Capital stock	0.183(0.012)**	0.150(0.024)**
Employment	0.180(0.010)**	0.135(0.017)**
Skill and capital intensity		
White-collar share of wage bill	0.001(0.001)	0.004(0.003)
White- to blue-collar worker ratio	0.002(0.001)+	0.001(0.003)
Capital stock per worker	0.010(0.006)+	0.023(0.017)
Productivity		
Labor productivity	0.001(0.004)	0.007(0.009)
TFP (Solow residual)	0.003(0.003)	0.004(0.007)
TFP (Ackerberg-Caves-Frazer)	0.005(0.003)	0.004(0.007)
Scope		
Number of products	0.145(0.008)**	0.102(0.015)**

Notes: (i) This table reports coefficients on log depreciated patent stock. All dependent variables, except white-collar share of wage bill and the white- to blue-collar wage ratio, are logged. (ii) The sample in column 1 includes all patentees. The sample in Column 1 is restricted to those firms that enter without a patent and switch to being patent-owning firm later in their life ("switchers"). (iii) The dependent variables are demeaned to control for industry-year effects. All regressions include firm fixed effects. (iv) The sample size in these regressions varies from about 60,000 in column 1 and 13,000 in column 2. (v) Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

TABLE 11: LINK WITH PATENT QUALITY

	Output (Differenced)		
	1	**2**	**3**
(1) Patent ownership dummy	0.331 (0.021)***	0.186 (0.031)***	0.360 (0.021)***
(2) Log number of patents in the first year	0.259 (0.021)***		
(3) Log number of forward citations to first-year patents		0.126 (0.014)***	
(4) Number of citations per patent for first-year patents			0.000 (0.002)
(1)*(2)	0.135 (0.017)***		
(1)*(3)		0.129 (0.014)***	
(1)*(4)			0.0153 (0.002)***
Number of observations	12,634	10,442	12,634

Notes: i) Patent ownership dummy is an indicator variable which equals 1 if the firm has applied for any patents up to and including the observation year. (ii) The dependent variable is log output less the mean of control group, matched on size, age, and industry in the year prior to patenting. (iii) The number of citations for any given patent is computed as the number of patents that cite the focal patent and were granted between the application year (of the focal patent) and 1999. (iv) Robust standard errors clustered at the industry level, * indicates significance at 10%, ** indicates significance at 5%. *** indicates significance at 1%.

APPENDIX 1: NAME-MATCHING RULES AND RELIABILITY CODES

A number of rules were applied to match the assignee names from the NBER patent dataset to the firm names from the SSEL. The rules varied by the "reliability code", with the degree of restrictions relaxed progressively. The first step in the matching process was obtaining a preliminary set of clusters using the SAS procedure DQMATCH, blocked on either the city and state or just the state (depending on the reliability code), with the sensitivity set at the highest level. This program creates clusters of firms (and assignees) with similar names, including those that sound similar. Furthermore, these clusters were based solely on the name, and not on which dataset the names came from. Hence, a cluster generated by this program could contain only firms from the SSEL (and no assignees from the patent dataset) or only assignees (and no firms from the SSEL) or a combination of both. All clusters with no assignees in them, and those with more than one assignee in them were dropped. The resulting set contained only clusters that had at least one SSEL firm and exactly one assignee. The names of the assignees and firms in these clusters were "cleaned" by applying rules depending on the reliability code. Any assignee and an establishment of a firm whose "cleaned" names matched **exactly** were treated as a match, with the assignee number being "bridged" to the Census Firm Number of the establishment.[1]

Reliability Codes 3, 4 and 43
All matches in these reliability levels are set to 1 if the (a) names of an assignee and the name of any of a firm's establishments (within a cluster identified in the preliminary DQMATCH procedure) matched as per the rules below (the rules vary by the code), and (b) the city and state of any of the firm's establishments in the SSEL matched the city and state of any of the assignee's inventors. Reliability code 3 has the most restrictive rules among these while reliability code 43 has the least restrictive rules among these codes.

Reliability Codes 5 and 53
All matches in these reliability levels are set to 1 if the (a) names of an assignee and the name of any of a firm's establishments (within a cluster identified in the preliminary DQMATCH procedure) matched as per the rules below (the rules vary by the code), and (b) the state of any of the firm's establishments in the SSEL matched the city and state of any of the assignee's inventors. Reliability code 5 has the more restrictive rules of these two codes. Since the matches in these codes use a state-level blocking, they are could have more matching errors than those that use both city and state to block the matches.

Name cleaning rules (Reliability Codes 3 and 5)
(a) Suffixes INC/INC./INCORPORATED/INCORPORATED./INCORP/INCORP/ CO INC. were assumed to be identical;

[1] Throughout, any assignees matched to exactly one establishment were identified separately using a dummy variable to enable any future analyses that may potentially require an establishment-level match.

(b) Suffixes CORP/CORP./CORPORATION/CORPORATION. were assumed to be identical;
(c) Suffixes COMPANY/COMPANY./CO./CO were assumed to be identical;
(d) Suffixes LIMITED./LIMITED/LTD./LTD were assumed to be identical;
(e) The word ASSOCIATION was assumed to be the same as the word ASSN;
(f) The words MANUFACTURING and INTERNATIONAL were assumed to be the same as the words MFG and INTL respectively;
(g) If the first word in a name was THE, that word was dropped;
(h) The symbols ",", "%", "(" and ")" were all dropped;
(i) The symbols "&" and "+" were treated as the word AND.

Name cleaning rule (Reliability Code 4)
In addition to the cleaning rules for reliability code 3, suffixes INC, CORP, CO, LTD, LLC and PLC, and the symbol "-" between two parts of the name were dropped. Hence, within this code, for instance, ABC INC in Sacramento, CA would be considered the same as ABC CORP in Sacramento, CA.

Name cleaning rules (Reliability Code 43)
In addition to the rules in reliability code 3, the following rules were applied to obtain matches under reliability code 43.
(a) Spaces between names were dropped as was the symbols "-" and "/" between two parts of the name.
(b) The letter "S" was dropped from the names (to identify names that differ only by the letter "S", mostly in the case of plurals)
(c) The word AND was dropped
(d) Drop the last words of a name if they were either OF or FSC.

Name cleaning rules (Reliability Code 53)
The rules are identical to reliability code 43, except that they are applied to clusters obtained using a state blocking instead of a city and state blocking in code 43.

A flow chart of the matching process is provided in the following figure. Table A.1 provides additional descriptive statistics on these matches.

Figure A.1: Flowchart of matching procedure

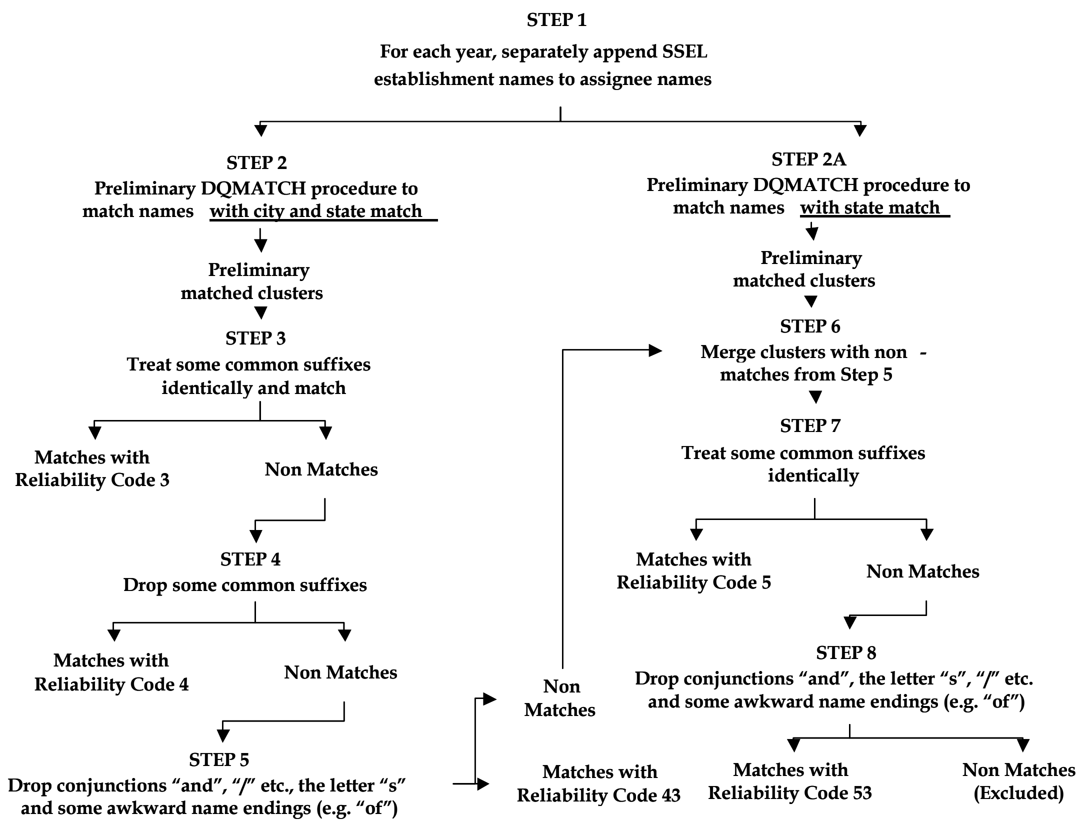


TABLE A.1: DESCRIPTIVE STATISTICS BY RELIABILITY CODE

(Contemporaneous Match Only)

Reliability Code	Number of Patents	Number of Assignee-years	Number of patents per assignee-year	Average number of citations to patents (per year)
3	510,036 (81%)	74,460 (61%)	6.85	39.55
4	40,116 (6%)	8,114 (7%)	4.94	29.94
43	8,826 (1%)	3,383 (3%)	2.61	13.43
5	64,366 (10%)	32,503 (27%)	1.98	10.78
53	5,755 (1%)	2,944 (2%)	1.95	11.14

APPENDIX 2: DEFINITIONS OF VARIABLES

Output: For any year before 1996, real Output is defined as the sum of shipments (deflated using the industry shipment deflator in the CES-NBER database developed by Bartelsman, Becker, and Gray, 2000) and the difference between year-beginning and year-ending deflated work in process and deflated finished goods inventories (the year-beginning inventory is deflated using the previous year's industry shipment deflator and year-end inventory deflated using the current year deflator). Following the LRD documentation, output for industries 2032, 2033, 2035, 2037, 2038, 2085, 2091, 2111, 2121, 2131 and 2141 was defined to be the sum of deflated shipments and the difference between year-beginning and year-ending deflated work in process inventories. Similarly, for industry 3731, output is defined to be the deflated shipments. For years including and after 1996, due to the unavailability of inventory data, output is simply defined to be the deflated shipments.

Value added: Real Value Added is defined as the difference between real output and real materials. Real materials are defined as the sum of deflated cost of material purchases, external contract work, fuel and electricity. For years before 1997, fuel and electricity are deflated using the energy deflator and the others using a materials deflator. Starting 1997, a single materials deflator is used.

Capital stock: This study uses the perpetual inventory approach to compute real capital stock. Separate stocks are computed for buildings (or structures) and machinery. Real capital stock (kt) in any given year, say for machinery, is computed as kt=(1-d)kt-1+It-1 + Rt where d is an industry-year specific depreciation rate for machinery, I is the capital investment in machinery (deflated by an industry-year specific investment deflator for the year t-1) and R is the capitalized value of capital equipment rentals. Depreciation rates were obtained from Chiang (2004) who bases his computations on BEA estimates, and Davis, Schuh and Haltiwanger (1996). Capital investment in machinery also includes investments in all used capital equipment (irrespective of machinery v. buildings) since the relevant variables are not well populated. The capitalized value of capital equipment rentals is obtained by dividing the rental expenditure by industry-year specific equipment rental prices obtained from the same source as above. This computation is done separately for buildings and machinery. Since establishments are not necessarily observed in their first year of operation (see entry year definition below), following Bahk and Gort (1993), capital stock in the first year (initial capital stock) is defined to be the book value of the assets at the end of the year deflated by an industry-year specific capital equipment deflator. This is defined separately for buildings and machinery where the decomposition is available. In case such a decomposition of year-end asset values is not available (1973,1988-91 and 1993-1997), all initial capital stock is assigned to machinery. If an establishment is not observed every year, following Olley-Pakes (1996), gross investment is imputed linearly (I^I_t=0.5*(I^O_t+I^O_{t-k})*(k-1), where I^I_t is the imputed investment for period t-k+1 to t-1, and I^O_t is the actual observed investment in period t).

Employment is measured as total number of employees reported by each firm.

White-collar share of wage bill is defined as the salary and benefits to non-production workers divided by total labor costs.

White- to blue-collar worker ratio is the number of non-production workers divided by the number of production workers.

Capital stock per worker is defined as Capital stock divided by Employment.

Labor productivity is defined as Output divided by Employment.

TFP (Solow residual) is defined as $TFP_{it}^{Solow} = y_{it} - \alpha_m m_{it} - \alpha_k k_{it} - \alpha_l l_{it}$, where y_{it} is the log of Output of firm i in year t, m is log real Materials (see definition of Value added above), k is log capital stock, and l is Employment. The elasticities α_m, α_k and α_l are defined equal to the material share, capital share and labor share of total costs in the industry j to which firm i belongs.

TFP (Ackerberg-Caves-Frazer) is defined using the methodology outlined in Ackerberg, Caves and Frazer (2005), who propose a methodology to control for the endogeneity of inputs which also addresses some of the shortcomings in the methodologies proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). For detailed discussion of the methodology used, refer Balasubramanian (2007).

Number of products is defined as the number of distinct 7 digit product codes reported by the firm.

Depreciated patent stock is defined as the total of depreciated patents, with each patent depreciated at an annual rate of 15 % (following Hall, Jaffe and Trajtenberg, 2005).

Industry: The study adopts 1987 SIC codes as the basis of industry definition. For most observations before 1997, the "ind" variable provided in the ASM and CM files is used to define the establishment's primary industry of operation. However, some of the observations wrongly use the 1972 codes, and are corrected per Chiang (2004) who bases these corrections on Davis, Haltiwanger and Schuh (1996). The magnitude and impact of these corrections is very small, though. In addition, establishments with non-existent SIC codes (e.g. 3079) are dropped. Starting 1998, the ASM files use the NAICS code; hence, the "oind" variable, which provides the "old" industry code for that establishment, is used.