DISCUSSION PAPER SERIES

高麗大學校　經濟研究所

THE INSTITUTE OF ECONOMIC RESEARCH
KOREA UNIVERSITY

『Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors』

Jinyoung Kim, Sangjoon John Lee, and Gerald Marschke

Discussion Paper No. 07-08 (April 2007)

The Institute of Economic Research Korea University
Anam-dong, Sungbuk-ku,
Seoul, 136-701, Korea
Tel: (82-2) 3290-1632 Fax: (82-2) 928-4948

Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors*

Jinyoung Kim†, Sangjoon John Lee††, and Gerald Marschke†††

†Korea University
††Alfred University
†††University at Albany and IZA

April, 2007

Abstract

It has long been recognized that worker wages and possibly productivity are higher in large firms. Moreover, at least since Schumpeter (1942) economists have been interested in the relative efficiency of large firms in the research and development enterprise. This paper uses longitudinal worker-firm-matched data to examine the relationship between the productivity of workers specifically engaged in innovation and firm size in the pharmaceutical and semiconductor industries. In both industries, we find that inventors' productivity increases with firm size. This result holds across different specifications and even after controlling for inventors' experience, education, the quality of other inventors in the firm, and other firm characteristics. We find evidence in the pharmaceutical industry that this is partly accounted for by differences between how large and small firms organize R&D activities.

JEL Classification: O30, O32, O34, J21, J24

Key words: Patents; Innovation; Labor productivity; Research; Firm size

---

* Lee is grateful to the National Bureau of Economic Research for the NBER-Sloan Science and Engineering Workforce Project Fellowship. All errors are exclusively the responsibility of the authors. Comments welcome at Jinyoung Kim, Department of Economics, Korea University, Seoul, Korea, jinykim@korea.ac.kr; and Gerald Marschke, Department of Economics & Department of Public Administration and Policy, SUNY Albany, BA-110, Albany, NY 12222, marschke@albany.edu.

I. Introduction

It has long been recognized that worker wages and possibly productivity is higher in large firms. Moreover, at least since Schumpeter (1942) economists have been interested in the relative efficiency of large firms in the research and development enterprise. This paper exploits panel data on inventors in the pharmaceutical and semiconductor industries, two industries that are prolific generators of innovations and patents, to examine the relationship between firm size and the productivity of workers specifically engaged in innovation. We use patents and patent citations as measures of inventor productivity. We link the inventors to firms in these industries through U.S. patent records, and obtain additional information on both the inventors and their employers from secondary sources. We find that in both industries, inventors' productivity increases with firm size. This result holds across different specifications and even after controlling for inventors' experience, educational level, the quality of other inventors in the firm, and other firm characteristics.

This paper is organized as follows. The next section summarizes the literatures on the worker productivity-employer size relationship and on the R&D-firm size relationship. Section III describes our empirical method and the dataset we created for this project. Section IV describes our empirical results and Section V concludes.

II. Literature Review

Do scientists' productivity vary across firms, and if so, how? Schumpeter (1942), Panzar and Willig (1981), and Cohen and Klepper (1996), among others, have argued that a firm's size may be an important determinant of its productivity in R&D, generally. Large firms with substantial market share may have an advantage in R&D because monopoly power enables them

to capture the returns to innovation. Large firms may also have an advantage because of either scale economies (due to fixed costs to mounting an R&D operation) or because their size affords them wider access to external sources of financing. Large firms with many product lines may be better able to exploit unexpected innovations.

There is considerable empirical research examining the relation between the productivity of R&D in firms and firm size. Such research has examined how patent (or citation-weighted patent) yields from R&D activities (usually measured by R&D expenditures) vary with firm size. This empirical work has failed to demonstrate a consistent pattern of increasing returns in R&D. While some studies find an advantage to size in R&D (e.g., Cohen & Klepper, 1996), many studies fail to show a relationship between size and R&D productivity, and still others find a negative relationship between size and R&D productivity (e.g., Acs and Audretsch, 1991, Bound et al, 1984, and Hausman et al, 1984). See Symeonidis, 1996, for a survey of this literature.[1]

To our knowledge no studies in the economic literature have examined the relation between firm size and worker productivity in the R&D enterprise.[2][3] A substantial literature documents employer size-wage or productivity premia generically, however. Large firms pay workers a premium that is comparable to or even greater than the wage gaps observed between genders, among races, and between unionized and non-unionized workers (see, for example, Oi, 1983; Brown and Medoff, 1989; Davis and Haltiwanger, 1991; and Troske, 1999). Moreover,

[1] In examining the relation between size and R&D productivity, researchers face difficulties obtaining accurate measures of R&D expenditures. A number of scholars (e.g., see Cohen and Levin, 1989) have discussed the strengths and limitations of the various measures of R&D inputs (R&D expenditures, scientific employment) and outputs (patents, patent citations). Many believe that R&D expenditures are especially poorly measured for smaller firms because smaller firms often conduct R&D informally (they lack R&D departments and separate R&D staff). This may explain why many studies fail to show a relationship between size and R&D productivity and other studies even find that the patent yield from R&D expenditures falls with firm size.

[2] A body of research exists (much of it outside economics) on other determinants (especially academic) scientist's productivity. See Stephan's (1996) survey.

[3] Some studies have examined the determinants (without examining firm size) of scientist earnings, however (see, e.g., Lillard and Weiss, 1979).

workers in large firms appear to be more productive. Many hypotheses have been put forward to explain the size-wage/productivity premium. For example, Idson and Oi (1999), Dunne and Schmitz (1992), and others have argued that labor productivity is higher in large firms because large firms adopt new technologies sooner and possess higher quality capital than small firms. In the context of R&D, this could imply that large firms possess newer and more sophisticated laboratories than small firms. Griliches (1970) and Hamermesh (1993) argue that the employer size-productivity premium is due to the complementarity between worker skill and physical capital. Large firms may employ inherently more able workers and impose higher work standards than small firms (see Idson and Oi).

III. Empirical Implementation

3.1 Model Specification

We use panel data on a sample of research scientists in industry to test whether scientists are more productive in large firms. We estimate the effect of firm size on the scientist's labor productivity in patenting using a Poisson-based maximum-likelihood regression model (Hausman, Hall, and Griliches, 1984). We favor a Poisson-based specification because the number of patents granted to a scientist in a particular year is a nonnegative count variable. We assume that the expected number of patents invented by a scientist, conditional on the scientist's characteristics and the characteristics of his/her firm, is

$$E(PAT_{it}) = \exp[\alpha + \beta_1 \ln(R\&D_{it}) + \beta_2 PHDEG_i + \beta_3 EXP_{it} + \beta_4 EXP_{it}^2 + \beta_5 PSTOCK\_5A_{it} + \beta_6 PSTOCK\_5B_{it} + \gamma X_{it}],$$

where PAT_{it} is the number of patents granted to scientist i that were applied for in year t, and $R\&D_{it}$ is year t R&D expenditures (deflated by the GNP deflator) of the firm to which scientist

i's patents are assigned. We use as a measure of firm size the magnitude of research expenditures instead of a more comprehensive firm size measure (such as firm sales) because it is likely a more relevant size factor in a scientist's productivity. To check the robustness of our result, however, we employ total sales (SALES) and the total number of employees (EMPLOYEE) as alternative size measures. Note also that we match a scientist with the firm to which the scientist's patents are assigned under the presumption that patents are assigned to inventors' employers. $PHDEG_i$ is a binary variable for those scientists with a Ph.D. degree, and EXP_{it} is the number of years elapsed at year t since scientist i had first been named to a patent. Following the Mincerian earnings regression studies, the two variables, EXP_{it} and EXP_{it}^2, are included to capture the scientist's experience in research. $PSTOCK\_5A_{it}$ and $PSTOCK\_5B_{it}$ also capture aspects of experience. $PSTOCK\_5A_{it}$ is the number of patents in the last 5 years (t-5 through t-1) on which the scientist is a named inventor. $PSTOCK\_5B_{it}$ is the number of patents accumulated until year t-5. $PSTOCK\_5A_{it}$ and $PSTOCK\_5B_{it}$ by reflecting preferences and ability capture the inclination to patent, but also they capture the experience and knowledge upon which scientist i can draw in his/her t period inventive activities. We anticipate finding a positive relationship between the likelihood of patenting at t and $PSTOCK\_5A_{it}$, and $PSTOCK\_5B_{it}$. Because recent experience and learning is likely more relevant, however, we expect a stronger relationship between patenting and $PSTOCK\_5A_{it}$. We use these two variables in linear form since they take non-negative values.

X_{it} is a vector of the (time-varying) characteristics of the firm to which scientist i's patents are assigned, all in logarithmic form. The vector includes the capital-labor ratio (K/L), the share of Ph.D. degree holders among all patenting inventors in the firm (PHD/INV), the mean experience of the patenting inventors (MEXP), firm age (FIRMAGE), and the number of

business lines (NSIC). To control for the quality of patents, we include the projected average number of citing patents per patent (MCITE) as an additional regressor in one specification. In an attempt to adjust the scientist's output measure for the number of collaborators, another specification includes the average number of inventors named on the scientist's patents in year t (COINVENT) as an additional regressor (we explain how we constructed these variables in the following section).

Because missing scientist-specific factors may affect the scientist's productivity in patenting, we estimate the models with scientist-specific random effects.

Our data include instances where in a given year a scientist is named as inventor on the patents assigned to multiple firms. These instances can take place, for example, when a scientist changes employers or when he/she is affiliated with one or more firms but not as an employee. In our analyses we exclude observations of this kind; that is, we exclude observations in which the scientist is matched to multiple firms in one year.[4]

3.2 Data and Variables Used

Our data are taken from six sources: (1) Patent Bibliographic data (Patents BIB) released by the U.S. Patent and Trademark Office (USPTO) that contain bibliographic information for U.S. utility patents issued from 1969 to 2002; (2) the Compact D/SEC database which contains firm information taken primarily from 10-K reports filed with the Securities and Exchange Commission; (3) the Standard & Poor's Annual Guide to Stocks-Directory of Obsolete Securities

[4] This exclusion drops the number of observations by 4.4% in pharmaceutical and 4.6% in semiconductor industry. In an alternative specification, omitted from this draft, we include such observations, assuming that the inventor is employed sequentially to each of the firms to which he/she is matched. For these observations, we assume the employment duration with any one firm is proportional to the ratio of patents assigned to that firm to patents assigned to all firms in that year. This alternative specification, however, leads to qualitatively similar results as we report in this paper.

which include a history of firm ownership changes due to mergers and acquisitions, bankruptcy, dissolution, and name changes; (4) the NBER Patent-Citations data collected by Hall, Jaffe and Trajtenberg (2001) which contain all citations made by patents granted in 1975-1999; (5) the Thomas Register data which report the firm's founding year, and finally (6) the ProQuest Digital Dissertation Abstracts database which contains information on the date, the field, and the type of degree for degree holders We match these data to the inventors in the Patents BIB data by scientists' names.

As the first step for merging these datasets, we choose all firms whose primary SIC is 2834 (pharmaceutical preparation) or 3674 (semiconductor and related devices) in the Compact D/SEC data. We selected these two industries for our study because the firms in these industries are active in patenting and produce homogenous products relative to other industries. Because patents are typically assigned to the firm that employs the inventors, we identify the inventors' employers in the Patents BIB data by patent assignees.

Because parent firms sometime patent under their own names and at other times under the names of their subsidiaries, however, merging the Patents BIB data with firm-level data in the Compact D/SEC data is not straightforward. Mergers and acquisitions at both the parent firm and subsidiary levels, common in these two industries during the 1990s, and name changes further complicate linking the patent to firm-level data. (The USPTO does not maintain a unique identifier for each patenting assignee at the parent firm level nor does it track assignee name changes.) Thus, to use the firm-level information available in the Compact D/SEC data, the names of parent firms and their subsidiaries and the ownership of firms must be tracked over the entire period of the study.

We use the S&P data to identify whether each assignee in our Patents BIB extract was a stand-alone firm. In some cases, due to a merger or acquisition, for example, the assignee was actually part of another firm at the time of the patent application. In these cases, we substituted this parent firm's name for the name given as the assignee. We then use the subsidiary data in the Compact D/SEC data to track changes in the parent for each of the firms (with the corrected names) in our Patents BIB extract. We thus assign our own firm identifier accordingly after tracking the histories of M&A's as well as of name changes of each firm. For example, if a firm is acquired, we keep separate data on the firm through the year prior to acquisition, because the acquiring firm reports consolidated information and because the patents applied for by the target firm after acquisition should be linked to the acquiring firm. The firm that acquires the target firm retains the same firm identifier. If firms are merged, we keep their observations up to the year prior to merger and assign the newly merged firm a new firm identifier. If a firm changes its name, it retains the same identifier. If a subsidiary's ownership changes, the subsidiary's identifier becomes the identifier of the new parent, from the date of the change forward.

After merging the Patents BIB data with the firm-level information in the Compact D/SEC data, we then link the patent inventors to the firms in the Compact D/SEC data by the final firm name to produce a data set on inventors and patents that includes firm-level data (e.g., R&D expenditures, sales, and number of employment) on the patents' assignees.

We obtained the inventors' educational backgrounds (degree types, dates, and fields for those who earned masters or doctoral degrees) from the ProQuest database, and linked this information to the patent and firm data by inventor name. Based on the list of those who earned

degrees in all natural science and engineering fields between 1945-2002, the inventors are matched to those on the list by their last, first and middle names.[5]

Information on all citations is from the NBER Patent-Citations data collected by Hall, Jaffe and Trajtenberg (2001). In these data each citing patent that was granted between 1975 and 1999 is matched to all patents cited by the patent. According to Hall et al., 50 percent of all citations are made to patents at least 10 years older than the citing patent, and 5% of citations refer to patents that are at least 50 years older than the citing one. Thus, we cannot observe the total number of citations for most of the patents in our data. We construct the projected number of all citing patents for each patent in our data as follows: Based on the average number of citing patents per cited patent in each ensuing year since application year for pharmaceutical and semiconductor patents in the USPTO data, we count what percentage of all citing patents cites a patent, for example, in the first 3 years after the cited patent's application year, and then estimate the total number of citing patents for a patent for which our data are censored 3 years after its application year.[6] We then calculate the projected average number of citing patents for scientist i in year t.

Definitions and summary statistics of variables used in our analysis are reported in table 1.

[5] Prior to the matching, we modified persons' names in both datasets by converting all lower case letters to upper case letters, deleted all non-alphanumeric characters, such as commas and hyphens. We noticed that inventors' middle names are sometimes reported inconsistently. That is, sometimes middle names are spelled out, sometimes only their initials are included, and at other times no middle name or initial is included. To achieve more accurate matching of inventors, only the initial was taken as middle name and then inventors with unique last, first and middle name were given a unique identifier. See Kim et al. (2004) for the detailed description of our data matching procedure.

[6] We choose patents in subcategories of 14, 19. 31-33 in in the USPTO classification for pharmaceutical industry and subcategories of 21, 22, 24, 41, 46 for semiconductor industry.

IV. Empirical Findings

We estimate the association between the number of an inventor's patent grants applied for in year t (PAT) and his/her employer's R&D expenditures in year t (R&D), our measure of the size of the firm's research enterprise. Respectively, tables 2 and 3 report separate estimation results for the pharmaceutical and semiconductor industries.[7] Our use of contemporaneous R&D, as opposed to lagged R&D, follows the extensive literature estimating patent production functions (e.g., Hall, Grilliches, and Hausman, 1986). Evidence suggests that R&D activities and innovations occur somewhat simultaneously. Moreover, if a firm attempts to patent an innovation, it files the application while the innovation is being developed or very shortly afterwards (Hall et al.).

Our base model (model 1) includes as regressors R&D in log, our measure of the size of the enterprise; additional firm-level regressors (X) to isolate and differentiate the effect of firm size from those of other firm characteristics; and the individual-level characteristics education, years of experience, and the two stock measures of past patenting. Following the Mincerian earnings regression studies, we employ both a linear and a quadratic experience term.[8] PHDEG is a binary variable indicating a holder of a Ph.D. degree from a university in a developed country. We expect that the majority of the inventors in our data set have at least a college education, and that those with PHDEG=0 have either a bachelor's or a master's degree, and sometimes a Ph.D. from a university in a developing country.

[7] Scientists appear in our data only if they appear on a patent. Presumably, many scientists in firms' R&D programs do not appear as inventors on patents in a given year, even though they are engaged in research. Our dependent variable is thus left-truncated at zero. In all Poisson-based estimations reported in tables 2 and 3, we adjust the likelihood for truncation.

[8] Education and years of experience are frequently shown to be significant in worker-level productivity and earnings regressions (e.g. Mincer, 1974, Card, 1999).

Models 2 and 3 use as an alternative measure for firm size, SALES and EMPLOYEE, respectively. Model 4 re-estimates the base specification assuming a fixed instead of a random effect. Often more than one scientist is named in a patent's inventor field. In such cases, perhaps, a single scientist's contribution or output is smaller than for patents when the scientist is named as sole inventor. Our variable COINVENT is our attempt to adjust the scientist's output measure for the number of collaborators. Model 5 includes the log of COINVENT as an additional regressor.

Our empirical exercise is designed to measure the effect of the size of the R&D enterprise on real output, which patents proxy. The economic values of the innovations underlying patents, however, vary considerably from patent to patent. Moreover, the economic value of many patents is close to zero. Evidence exists that a patent's citation by a subsequent patent—each patent documents the "prior art" upon which the new innovation builds—is an indicator not only of the importance of the underlying innovation but its economic value as well (see Trajtenberg, 1990). We created the variable MCITE, the average number of citations received by scientist i's patents in year t, to capture the quality of the scientist's patents (see section III for a description of this variable). Model 6 includes MCITE as an additional regressor. Thus model 6 generates an estimate of the effect of size on patent counts holding constant patent quality.

In models 1 through 6, the specifications include scientist-specific random or fixed effects. An alternative specification might include a separate firm-specific effect, to take into account unobserved firm characteristics that influence scientist productivity. To address this, we created a sub sample by omitting all scientists from our sample who were named inventors on more than one assignee's patent. We thus deleted from our sample scientists who move and patent for their different employers. Model 7 re-estimates the base specification using this sample of non-movers. In model 7, therefore, the random effects take care of both the scientist-specific and firm-specific

effects. Moreover, we mitigate the problem of potential name mismatches by deleting these observations in this model.

Controlling for the individual characteristics, the results of model 1 in both industries indicate that inventor productivity increases with the size of the R&D enterprise. The effect of firm size (R&D) on patent productivity (PAT) is positive and statistically significant in all models in Table 3 and most models in Table 2. (Note that in the Poisson specification the estimated coefficients for the log-transformed regressors have an elasticity interpretation.) Controlling for patent citations (see model 6) does not change the estimated size effects. The two other measures for firm size exhibit qualitatively the same effect on PAT (models 2 and 3). Note also that the magnitude of the firm size effect on patenting is generally similar in both the semiconductor industry and the pharmaceutical industry. These results support the finding in the labor literature that worker wages and productivity are higher in large firms,[9] but contrasts with findings elsewhere that small firms have higher patent-R&D ratios than large firms. In fact, Kim et al. (2003), using the firm-level variables constructed from the same databases as those in this paper, show that the estimated relationship between firm size and patenting productivity of R&D expenditures at the firm level is significantly negative.

Controlling for the number of collaborators (model 5) does not change the estimated size effect in the semiconductor regressions. In the pharmaceutical regressions, however, controlling for the number of collaborators does reduce the magnitude of the coefficient on ln(R&D). Adding the regressor ln(COINVENT) also makes the coefficient estimate only marginally significant. This suggests that the patent-firm size effect in the pharmaceutical industry is partly due to the way the R&D enterprise is organized. Table 4 confirms that the average number of

[9] Our elasticity estimates are smaller than have been reported for manufacturing workers. The elasticity estimates reported by Idson and Oi, for example, range between .09 and .18.

collaborators on a patent rises with firm size. This relationship is especially pronounced in the pharmaceutical industry, where the number of inventors per patent is 25 percent greater in the top size quartile compared to the bottom quartile.

Note that the magnitude of the coefficient estimates on ln(R&D) in the estimation using non-movers only (model 7) is similar to the magnitude using the full sample (model 1). In the pharmaceutical industry regression, however, the t statistic falls from 2.29 to 1.82.

Human capital theory predicts that higher education is associated with higher productivity and that labor productivity has an inverted U-shaped relationship with experience. Both predictions are strongly confirmed in tables 2 and 3. PHDEG exhibits a significant and positive effect on PAT, and EXP has an inverted U-shape relationship with PAT.[10] According to the estimated coefficients in model 1 of tables 2 and 3, the peak in PAT is reached at EXP=13 years and 11.9 years in pharmaceutical and semiconductor industry, respectively. The coefficient estimates on PSTOCK_5A and PSTOCK_5B are generally both positive and significant, as predicted. (The coefficient estimate on PSTOCK_5B in the fixed effects specification in Table 2 is statistically insignificant). Also, recent inventions are more strongly correlated with patent productivity than earlier ones.

We include the capital labor ratio as a regressor because given R&D expenditures a highly capitalized firm may have a stronger incentive to patent than less capitalized firms. A patent infringement lawsuit that leads to production stoppage will be more destructive for a firm that has made a large capital investment in a state-of-the-art physical plant. Such vulnerability may encourage the firm to develop a diverse portfolio of patents that it can use as a bargaining chip to ward off infringement suits (Cohen et al., 2000; Parr and Sullivan, 1996). In addition, we

expect that the capital intensity in the R&D division will be positively correlated with the capital intensity in the entire firm. Therefore, a higher K/L in the entire firm may raise labor productivity in patenting of an inventor due to capital-skill complementarity. The results in both tables show support for this prediction. In tables 2 and 3, K/L in log exhibits a significant and positive effect on PAT in all models.[11]

A higher percentage of Ph.D. degree holders among inventors may raise the labor productivity of an inventor in the same firm due to a positive spillover effect. On the other hand, firms with a higher concentration of Ph.D. degree holders may be engaged in innovations with higher economic values and produce a smaller number of patents. The net effect of the proportion of Ph.D. degree holders on labor productivity is thus theoretically ambiguous. In the pharmaceutical regressions, our results show that PHD/INV has a positive and generally significant effect on PAT. (In the fixed effects estimation in the pharmaceutical industry regressions, model 4, Table 2, the coefficient estimate is statistically insignificant.) On the other hand, in the semiconductor industry, corresponding coefficient estimate is either insignificant or significantly negative.

In the pharmaceutical industry regressions, the coefficient estimate on the median experience of patenting inventors in a firm, MEXP, is statistically significant only in the fixed effects specification. In the semiconductor industry regressions, this coefficient estimate is statistically significant only in the model that contains the citation measure, and in that model,

---

[10] Note, however, that because we are not controlling for attrition, which likely occurs selectively, the inverted U-shaped relationship is due not only to changes in the productivity of a given cohort of scientist, but to the changing composition of the cohort as well.

[11] To further control for the capital intensity in research, we have estimated a specification with the log of R&D expenditures per (patenting) inventor as an additional regressor. Ideally, such a measure captures the R&D resources (laboratory capital and colleagues, for example) accessible to the scientist. Because we only observe the subset of inventors who patent, however, R&D/INV is a noisy measure of this characteristic at best, and is likely biased. The size effect on scientist productivity is still significantly pronounced in this specification.

the coefficient estimate is negative. Thus, we find no clear evidence of positive productivity spillovers across inventors within a firm.

Tables 2 and 3 show that the age of a firm is negatively related to the patent productivity of the firm's inventors (except in the fixed effects specification for the semiconductor industry). Two possible reasons may be that (1) older firms carry out larger scale projects and produce a smaller number of patents with higher economic value, or (2) older firms have exhausted new ideas for innovations and produce fewer patents.

Our results indicate that in both industries, the number of business lines in a firm, measured by the number of secondary SIC's classified to the firm (NSIC), has a negative and generally significant effect on PAT. This evidence does not support the presence of economies of scope in scientific labor productivity. The evidence may reflect instead varying mixes of the technologies researched across firms of varying NSIC. That is, scientists in firms with multiple lines may be working in fields where the economic value of patents is large, relative to the fields in which scientists in firms with few business lines.

V. Conclusion

Our findings can be summarized as follows. Using patents as our measure of a scientist's output, we find that labor productivity of scientists rises with firm size, whether size is measured in R&D expenditures, sales or employees. Our finding that patent productivity rises with firm size even after controlling for measures of the ability of scientists in the firm (including scientist-specific fixed effects), suggest that productivity advantages enjoyed by large firms is not simply due to large firms' ability to hire and retain high quality researchers. Our results are robust across different specifications, including specifications that control for unobserved firm and

scientist heterogeneity, and arise in both the pharmaceutical and semiconductor industry analyses. This is especially interesting because at the firm level in these two industries, the patent yield per R&D spent varies *inversely* with firm size.

An alternative interpretation supported by the pharmaceutical industry data is not that scientists in large firms are more productive, but that the scientific enterprise is organized differently in large firms. For example, it is possible that in large firms or in firms with large R&D enterprises, a scientist plays a smaller role in any single R&D project. In a large firm, due to the opportunity for specialization, a single scientist in a given period performs smaller tasks on any one project but contributes to more projects than his/her counterpart in a small firm. Consistent with this story, we find that the scientist-firm size productivity effect in the pharmaceutical industry disappears when we control for the number of collaborators.

References

Acs, Z.J. and Audretsch, D.B. "Innovation, Market Structure, and Firm Size," *The Review of Economics and Statistics*, Vol. 69, No. 4, pp. 567-574, Nov. 1987.

Acs, Z.J. and Audretsch, D.B. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, 78, 678-90, 1988

Baldwin, W.L. and Scott, J.T. *Market Structure and Technological Change.* Chur, Switzerland: Harwood Academic Publishers, 1987

Bound, John, Cummings, Clint, Griliches, Zvi, Hall, Bronwyn H., and Jaffe Adam. "Who Does R & D and Who Patents?" in *R&D, Patents, and Productivity*, ed. Z. Griliches, NBER, 1984.

Brown, Charles and Medoff, James. "The Employer-Size-Wage Effect." *Journal of Political Economy.* Vol. 97, No. 5, pp. 1027-59, October 1989.

Card, David, "The Causal Effect of Education on Earnings," in Handbook of Labor Economics, ed. Orley Ashenfelter and David Card, Elsevier, New York, 1999.

Chandler, A. D. Scale and Scope*: The Dynamics of Industrial Capitalism*, Cambridge, Mass: MIT Press, 1990.

Cohen, W.M. and Klepper, S. "A Reprise of Size and R & D," *The Economic Journal*, Vol. 106, No. 437, pp. 925-951, July 1996.

Cohen, W.M. and Levin, R.C. "Empirical Studies of Innovation and Market Structure," in *Handbook of Industrial Organization,* eds R. Schmalensee and R.D. Willig, New York: North-Holland, 1989

Comanor, W.S. "Research and Technical Change in the Pharmaceutical Industry," *Review of Economics and Statistics*, Vol. 47, pp. 182-90, 1965

Davis, Steven J. and Haltiwanger, John. "Wage Dispersion Between and Within Manufacturing Plants, 1963-1986." In *Brookings Papers on Economic Activity*, pp. 115-200, 1991.

Dunne, Timothy and Schmitz, James A. "Wages, Employment Structure and Employer Size-Wage Premia: Their Relatinship to Advanced-Technology Usage at U.S. Manufacturing Establishments." Discussion paper no. 92-15 Washington, D.C.: Bureau of the Census, Center for Economic Studies (December 1992).

Freeman, C. The Economics of Indsutrial Innovation, 2d. ed. Cambridge, Mass: MIT Press, 1982

Graves, S.B. and Langowitz, N. "Innovative Productivity and Returns to Scale in the Pharmaceutical Industry," *Strategic Management Journal*, Vol. 14, pp. 593-605, 1993

Griliches, Zvi. "Notes on the Role of Education in production Functions and Growth Accounting." In W. Lee Hansen (ed.), Education, Income, and Human Capital (New York: NBER, 1970).

Hamermesh, Daniel S. *Labor Demand.* (Princeton, NJ: Princeton University press, 1993.)

Hall, B. H., A. B. Jaffe, and M. Tratjenberg "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498, 2001.

Hausman, Jerry A., Hall, Bronwyn H., and Griliches, Zvi. "Econometric Models for Count Data with an Application to the Patents R and D Relationship," *Econometrica* 52, 1984.

Henderson, Rebecca and Cockburn, Ian, "Scale, scope, and spillovers: the determinants of research productivity in drug discovery," *RAND Journal of Economics*, 27(1), pp. 32-59, Spring 1996.

Idson, Todd L. and Oi, Walter Y. "Workers are More Productive in Large Firms." *American Economic Review.* Vol 89, No.2, pp. 104-108, May 1999.

Jaffe, Adam B., Trajtenberg, Manuel, and Henderson, Rebecca. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, Vol. 108, No. 3, pp. 577-598, August 1993

Jensen, E.J. "Research Expenditures and the Discovery of New Drugs," *Journal of Industrial Economics*, Vol. 36, pp. 83-95, 1987.

Kim, Jinyoung, Lee, Sangjoon, and Marschke, Gerald. "R&D Scientists in Pharmaceutical and Semiconductor Industries: Data Construction and Description," working paper, February 2004.

Kim, Jinyoung, Lee, Sangjoon, and Marschke, Gerald. "Relation of Firm Size to R&D and Innovative Output," October 2003.

Kleinknecht, A. "Measuring R & D in Small Firms: How Much are we Missing?" *The Journal of Industrial Economics*, Vol. 36, No. 2, pp. 253-256, Dec. 1987.

Lillard, Lee A. and Weiss, Yoram. "Components of Variation in Panel Earnings Data: American Scientists 1960-70," *Econometrica* , Vol. 47, No. 2, pp. 437-454, Mar. 1979.

Mincer, Jacob, *Schooling, Experience, and Earnings*, New York, Columbia University Press, 1974.

Oi, Walter Y. "The Fixed Employment Costs of Specialized labor." In Jack E. Triplett (ed.), *The Measurement of Labor Costs* (Chicago: University of Chicago Press for NBER, 1983).

Panzar, J.C. and Willig, R.D. "Economies of Scope (in Sustainability Analysis)," *The American Economic Review*, Vol. 71, No. 2, pp. 268-272, May, 1981.

Pavitt, K., Robson, M., and Townsend, J. "The Size Distribution of Innovative Firms in the U.K.: 1945-1983," *Journal of Industrial Economics*, V. 35, 297-316, 1987.

Scherer, F.M. "R & D and Declining Productivity Growth," *The American Economic Review*, Vol. 73, No. 2, pp. 215-218, May, 1983

Schwartzman, D. *Innovation in the Pharmaceutical Industry*. Baltimore: Johns Hopkins University Press, 1976

Schumpeter, J. *Capitalism, Socialism, and Democracy*. Harper and Row, New York, 1942.

Stephan, P. E. "The Economics of Science," *Journal of Economic Literature*, Vol. 34, Issue 3, pp. 1199-1235, Sep. 1996.

Trajtenberg, Manuel. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The Rand Journal of Economics.* Vol 21, No. 1, pp. 172-87, Spring 1990.

Troske, Kenneth. "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *The Review of Economics and Statistics.* Vol 81, No. 1, pp. 15-26, February 1999.

Vernon, J.M. and Gusen, P. "Technical Change and Firm Size: The Pharmaceutical Industry," *Review of Economics and Statistics*, Vol. 56, pp. 294-302, 1974

Table 1 Variable Definition and Sample Statistics

	Definition [Data Source]	Mean (Standard Deviation) Pharmaceutical	Semiconductor
PAT	Number of patents granted to a scientist by application year [Patents BIB]	1.587 (1.429)	2.024 (2.717)
R&D	Real R&D expenditures in 1996 constant dollars [Compact D/SEC]	9,433 (3,999)	6,144 (6,672)
SALES	Real sales volume in 1996 constant dollars [Compact D/SEC]	92,353 (47,126)	58,470 (71,545)
EMPLOYEE	Number of Employees [Compact D/SEC]	39,334 (13,921)	21,819 (19,584)
PHDEG	Binary variable for a Ph.D. degree holder [ProQuest]	0.303 (0.459)	0.128 (0.334)
EXP	Years elapsed since the inventor's first patent granted is applied [Patents BIB]	6.314 (6.175)	4.604 (5.152)
PSTOCK_5A	Number of patents in years t-5 through t-1 on which the scientist is a named inventor [Patents BIB]	2.573 (4.281)	2.429 (5.207)
PSTOCK_5B	Number of patents accumulated before year t-5 on which the scientist is a named inventor [Patents BIB]	2.204 (5.860)	0.879 (2.957)
K/L	Capital-labor ratio, or deflated plant and equipment over the number of employees [Compact D/SEC]	0.917 (0.382)	1.313 (1.541)
PHD/INV	Share of inventors who hold Ph.D. degrees [ProQuest, Patents BIB]	0.305 (0.070)	0.133 (0.067)
MEXP	Median experience of all inventors in a firm [Patents BIB]	6.928 (0.921)	5.021 (1.213)
FIRMAGE	Years elapsed since the founding year of a firm [Thomas Register]	92.372 (37.040)	28.603 (19.372)
NSIC	Number of secondary SIC's assigned to a firm [Compact D/SEC]	5.258 (1.902)	2.408 (1.611)
COINVENT	Average number of inventors named on the scientist's patents in year t [Patents BIB]	4.373 (2.921)	2.973 (1.854)
MCITE	Average projected number of citations per patent [Citation]	5.579 (11.032)	6.783 (12.354)

Table 2 Scientist Patent Productivity: Pharmaceutical

Poisson Model with Truncation and Random Effects

	(1) Base	(2) Sales	(3) Employees	(4) Fixed Effects
ln(R&D)	0.0269			0.0622
	2.29			3.20
ln(SALES)		0.0276		
		3.27		
ln(EMPLOYEE)			0.0326	
			3.50	
PHDEG	0.0730	0.0735	0.0746	0.0699
	6.10	6.16	6.29	2.13
EXP	0.0260	0.0262	0.0256	0.0474
	8.52	8.61	8.47	8.66
EXP^2	-0.0010	-0.0011	-0.0010	-0.0013
	-7.43	-7.51	-7.38	-4.80
PSTOCK_5A	0.0363	0.0365	0.0368	0.0141
	50.37	50.44	52.06	8.10
PSTOCK_5B	0.0029	0.0030	0.0029	0.0026
	3.75	3.74	3.75	1.07
ln(K/L)	0.1243	0.1329	0.1414	0.0719
	7.26	7.23	7.84	2.45
ln(PHD/INV)	0.0811	0.0801	0.0926	0.0634
	2.82	2.81	3.37	1.42
ln(MEXP)	0.0208	0.0040	0.0158	0.1767
	0.42	0.08	0.32	2.32
ln(FIRMAGE)	-0.0233	-0.0206	-0.0231	-0.0777
	-2.12	-2.00	-2.22	-3.44
ln(NSIC)	-0.0213	-0.0289	-0.0324	-0.1528
	-1.55	-2.02	-2.20	-6.63
Observation	20829	20994	21182	20829
Log Likelihood	-29729	-29921	-30178	-13561

Table 2 Scientist Patent Productivity: Pharmaceutical (Continued)

	Poisson Model with Truncation and Random Effects		
	(5) Co-inventors	(6) Citations	(7) Non movers
ln(R&D)	0.0206	0.0345	0.0281
	1.75	2.86	1.82
PHDEG	0.0750	0.0747	0.0910
	6.25	6.23	6.21
EXP	0.0262	0.0259	0.0370
	8.55	8.45	9.84
EXP^2	-0.0011	-0.0010	-0.0014
	-7.49	-7.33	-8.16
PSTOCK_5A	0.0359	0.0364	0.0357
	50.65	50.36	43.70
PSTOCK_5B	0.0030	0.0028	0.0013
	3.80	3.56	1.43
ln(K/L)	0.1159	0.1273	0.1240
	6.68	7.43	5.75
ln(PHD/INV)	0.0747	0.0833	0.0792
	2.59	2.88	2.02
ln(MEXP)	0.0386	0.0319	0.0266
	0.78	0.64	0.40
ln(FIRMAGE)	-0.0208	-0.0277	-0.0107
	-1.89	-2.52	-0.75
ln(NSIC)	-0.0191	-0.0297	-0.0312
	-1.38	-2.15	-1.60
ln(COINVENT)	0.0828		
	7.01		
MCITE		0.0060	
		6.09	
$MCITE^2$		-7.89E-05	
		-5.09	
Observation	20829	20829	14488
Log Likelihood	-29693	-29705	-20614

Note: Two rows for each variable report the coefficient and the ratio of the coefficient to its standard error. The coefficient estimates for constant terms are omitted from the table due to space considerations.

Table 3 Scientist Patent Productivity: Semiconductor

Poisson Model with Truncation and Random Effects

	(1) Base	(2) Sales	(3) Employees	(4) Fixed Effects
ln(R&D)	0.0265			0.0508
	4.05			3.76
ln(SALES)		0.0393		
		8.09		
ln(EMPLOYEE)			0.0431	
			9.26	
PHDEG	0.0881	0.1140	0.1133	0.1828
	5.94	10.78	10.71	4.03
EXP	0.0668	0.0506	0.0508	0.0995
	23.48	24.14	24.23	15.80
EXP^2	-0.0028	-0.0021	-0.0021	-0.0036
	-20.52	-20.96	-21.00	-11.85
PSTOCK_5A	0.0257	0.0277	0.0276	0.0075
	184.93	238.90	237.11	8.45
PSTOCK_5B	0.0066	0.0050	0.0049	0.0120
	5.95	6.07	6.10	4.29
ln(K/L)	0.0467	0.0210	0.0442	0.0180
	11.87	6.44	13.89	2.37
ln(PHD/INV)	-0.0349	-0.0567	-0.0588	0.0040
	-2.31	-4.48	-4.64	0.15
ln(MEXP)	-0.0432	0.0019	0.0184	0.0244
	-1.34	0.07	0.66	0.44
ln(FIRMAGE)	-0.0895	-0.1530	-0.1627	-0.0004
	-5.21	-14.03	-14.72	-0.01
ln(NSIC)	-0.0514	-0.0604	-0.0546	-0.0168
	-6.10	-8.68	-8.27	-0.88
Observation	17867	33294	33294	17867
Log Likelihood	-30557	-53618	-53606	-12315

Table 3 Scientist Patent Productivity: Semiconductor (Continued)

	Poisson Model with Truncation and Random Effects		
	(5) Co-inventors	(6) Citations	(7) Non movers
ln(R&D)	0.0254	0.0229	0.0295
	3.86	3.37	3.21
PHDEG	0.0891	0.0891	0.0892
	5.97	6.02	4.59
EXP	0.0675	0.0665	0.1185
	23.61	23.17	23.85
EXP^2	-0.0028	-0.0027	-0.0056
	-20.64	-20.32	-14.24
PSTOCK_5A	0.0255	0.0251	0.0204
	178.42	171.89	111.64
PSTOCK_5B	0.0067	0.0066	-0.0049
	5.93	5.96	-1.80
ln(K/L)	0.0455	0.0411	0.0405
	11.55	10.29	8.18
ln(PHD/INV)	-0.0443	-0.0316	-0.0177
	-2.90	-2.06	-0.85
ln(MEXP)	-0.0485	-0.1129	-0.0443
	-1.48	-3.08	-0.99
ln(FIRMAGE)	-0.0879	-0.0948	-0.1103
	-5.15	-5.46	-4.71
ln(NSIC)	-0.0508	-0.0466	-0.0471
	-5.93	-5.49	-4.68
ln(COINVENT)	0.1324		
	14.33		
MCITE		-0.0042	
		-4.12	
$MCITE^2$		-5.71E-06	
		-0.35	
Observation	17867	17867	13686
Log Likelihood	-30489	-30523	-22993

Note: Two rows for each variable report the coefficient and the ratio of the coefficient to its standard error. The coefficient estimates for constant terms are omitted from the table due to space considerations.

Table 4 Average Number of Co-Inventors per Patent by Firm Size

Firm R&D Size	Pharmaceutical Industry	Semiconductor Industry
Bottom 25%	3.926	2.980
25-50%	4.301	2.805
50-75%	4.411	2.875
75-100%	4.859	3.216