Hi-Stat

Discussion Paper Series

No.229

Estimating Production Functions with R&D Investment and Edogeneity

Young Gak Kim

December 2007

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences

A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

Estimating Production Functions with R&D Investment and Endogeneity

YoungGak Kim
Graduate School of Economics
Hitotsubashi University

December 2007

ABSTRACT

This study analyses the production function estimation when there is an unobservable idiosyncratic productivity shock and the series of the productivity shock follows a first-order endogenous Markov process which is controlled by R&D investment.

The production function approach, in general, suffers from endogeneity problems when there are determinants of production which are not observed by the econometrician but are observed by the manager of a firm. To control for this problem, recently developed econometric methods are applied to the production function estimation. The results show that there is a possibility that other estimation methods such as OLS estimation and fixed effect estimation underestimates the contribution of capital. The results also suggest that the rate of return to R&D varies considerably across industries and within an industry.

JEL Classification Numbers: D24, O32

1. Introduction

In the literature, there are two main approaches to the valuation of R&D, that is, the production function approach and the market value approach. In the production function approach, the production function is typically estimated in a parametric way with the R&D stock which is the accumulated value of R&D expenditure after depreciation. This approach was originated by Griliches (1981), and is often called the knowledge-capital model. In contrast, the market value approach is based on the theorem that under the assumptions of linear homogeneities of the production function and the adjustment cost function, the value of a firm is equivalent to the weighted sum of each asset of the firm.[1] By regressing the market value of a firm on the assets which the firm has, one can obtain the coefficients on the assets which show how the market values each asset. The market value approach also uses the R&D stock to gauge the rate of return of R&D.

Although these two approaches have a long history and have been used in many studies, both suffer from some common problems. Here, two of them are singled out.

The first is the endogeneity problem in the estimation. As Marschak and Andrews (1944) and many other scholars after them have pointed out, there is a possibility that the decisions that a firm makes depend on the productivity which is unobservable to the econometrician. If this is the case, OLS estimates are biased, no matter whether the production function or of market value approach are chosen.[2] A few solutions to this problem have been developed and used in the literature. Two of the earliest solutions are instrumental variables (IV) estimation and fixed effects estimation. For the last fifteen years, two new techniques have been developed. One of them is dynamic panel estimation.[3] The other is the technique developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003).

The second problem is related with how to construct the R&D stock data. In this literature, many studies calculate accumulated R&D expenditures (with appropriate depreciation) to obtain the R&D stock. However, this procedure requires the certainty assumption (i.e., all of the R&D expenditure is accumulated with 100 percent certainty) and assumptions with regard to the depreciation rate (i.e., R&D stock depreciates with a certain fixed rate). However, the first assumption ignores the uncertainty surrounding R&D, which is one of the most characteristic aspects of R&D investment. As for the second assumption, most studies simply assume an arbitrary rate of depreciation, which has been traditionally 15 percent. If one tries to use only flow information on R&D to estimate the production function, another strong assumption is required, namely that the ratio of R&D expenditure to R&D stock is very stable, which may not be the case in reality.

This paper focuses on addressing these two problems taking the production function approach.

[1] See Wildasin (1984).

[2] The market value estimation is basically equivalent to the investment function estimation which is well known to suffer from the endogeneity problem.

[3] See Arellano and Bover (1995) and Blundell and Bond (2000).

The estimations in this paper are designed for solving the endogeneity problem, and are based on the estimation model developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). Both these studies propose a similar structural estimation, but differ in which variable to use to proxy the firm-specific productivity shock. Their approaches have been adopted and extended in a large number of studies.[4] Yet, most of these studies do not explicitly include R&D activity as an input. As for this point, Buettner (2005) showed that endogenizing the productivity process and incorporating R&D expenditures into the dynamic investment model of Olley and Pakes is difficult. However, Doraszelski and Jaumandreu (2007) tried to solve this problem using labor input to proxy the productivity shock and obtained reasonable results. The estimation model of Doraszelski and Jaumandreu (2007) is closest to the model in this paper.

The estimation model in this paper uses only information on R&D expenditure, not R&D stock. Instead, R&D expenditure is assumed to contribute to the enhancement of the productivity. One of the advantages of this model is that there is no need for strong assumptions with regard to the R&D activity, such as assumptions of a fixed rate of depreciation and the linear and certain accumulation of knowledge.

This paper found that there are potential estimation biases and that the possible origins of the biases are unobservable productivity shocks (endogeneity problem) and ignoring the contribution of R&D activity. The biases are especially prominent in the estimates of the coefficient on capital. Besides, the relationships between returns on R&D investment and firm's characteristics are examined in detail.

The paper starts by describing the basic model in the next section. In Section 3, the strategy for controlling for the endogeneity is described in detail. Section 4 provides some explanation on the data used in this paper. In Section 5, the result of the basic estimation is reported and compared with the results of other estimation methods, while Section 6 focuses on the relationship between productivity and R&D expenditure.

2. The optimal behavior of a firm

This section describes the basic model used in this paper. The model considers the optimal behavior of a firm in a general setting. At the beginning of each period, the firm makes four input decisions, that is, how much intermediate input to use for the production of that period, how much to invest as capital for the production of the next period, how much to expend on R&D activity to enhance the productivity of next period, and how many people to employ for the production of next period. These assumptions mean that input decisions with regard to capital, labor, and R&D should be made one year ahead. In most studies, labor input is considered as a variable input. In reality, however, and

[4] According to Ackerberg, Caves, and Frazer (2006), these two studies are cited directly in more than 800 papers.

especially in Japan, labor mobility is relatively low, meaning that labor is more akin to a fixed input. Most listed companies decide their employment levels at least one year ahead.

Given this situation, the only variable input is intermediate input. Investment in tangible capital, R&D activity, and employment decisions may involve adjustment costs. An optimizing firm maximizes the discounted present value of future profits.

The production function of such a firm is

$$Y_{it} = F(A_t, K_{it}, L_{it}, M_{it}, \omega_{it}, \varepsilon_{it}), \tag{1}$$

where K_{it}, L_{it}, and M_{it} are the capital input, labor input, and intermediate input respectively. A_t is the common technology level at time t. Both ω_{it} and ε_{it} are productivity shocks experienced by the firm, but differ in the sense that the former are observable to the manager of the firm but not to the econometrician whereas the latter are unobservable to and unpredictable by both the manager and the econometrician.

Each period, the manager of the firm decides the level of production (Y_{it}), of intermediate inputs (M_{it}), of investment in capital for production in the next period (I_{it}), of expenditure on R&D activity for production in the next period (R_{it}), and of additional employment for production in the next period (E_{it}) after observing the firm's productivity (ω_{it}). These assumptions mean that it takes a full period for new investment in capital to be ordered and installed and additional employees to be hired and trained and to be ready for production. New investment and new employees add to the future capital stock and labor force respectively in a deterministic way:

$$K_{it+1} = (1-\delta)K_{it} + I_{it} \tag{2}$$

$$L_{it+1} = L_{it} + E_{it} \tag{3}$$

Investment in R&D activity does not show up in the production function because it is assumed here that the firm invests in R&D activity to enhance the productivity level. The productivity shock in the next period is assumed to be a function of the productivity level and the R&D expenditure in the current period:[5]

[5] In this paper, it is assumed that the productivity shock is the only unobservable variable. R&D activity contributes to production only through the productivity process. An alternative approach would be to assume there are two unobservable variables: a productivity shock and the knowledge capital generated by past R&D investment. However, this approach would require much more

$$\omega_{it+1} = G(\omega_{it}, R_{it}) \tag{4}$$

Function G is assumed to include a stochastic term and to be expectable by the manager. In much of the literature on knowledge capital, the functional form of G is assumed to be deterministic and linear. Here, however, this assumption regarding the functional form is not imposed. Olley and Pakes (1996) and Levinsohn and Petrin (2003) assume an exogenous first-order Markov process for equation (4) which excludes R_{it}. This issue will be discussed in greater detail below.

The optimization problem the firm faces can be expressed by the following Bellman equation:

$$\begin{aligned} V(K_{it}, L_{it}, \omega_{it}) = \max_{I_{it}, E_{it}, R_{it}} \{ & P_t^Y F(A_t, K_{it}, L_{it}, M_{it}, \omega_{it}, \varepsilon_{it}) \\ & - P_t^K \{K_{it} + C^K(I_{it}, K_{it})\} - P_t^L \{L_{it} + C^L(E_{it}, L_{it})\} - P_t^R C^R(R_{it}, \omega_{it}) - P_t^M M_{it} \\ & + \frac{1}{1+\rho} \mathrm{E}\left[V(K_{it+1}, L_{it+1}, \omega_{it+1}) \mid K_{it}, L_{it}, \omega_{it}, I_{it}, E_{it}, R_{it}\right]\} \end{aligned} \tag{5}$$

C^K, C^L, and C^R represent the adjustment cost of capital investment (investment in tangible capital, I_{it}), labor (new employment, E_{it}), and R&D activity (R&D expenditure, R_{it}). P^Y is the output price and P^K, P^L, P^M, and P^R are the factor prices of capital, labor, intermediate input, and R&D. ρ is the discount rate. Adjustments in intermediate inputs do not incur any costs, so that the firm can flexibly change intermediate input levels as required.

The optimizing behavior can be described with a set of first order necessary conditions. The condition with respect to intermediate inputs is simplest:

$$\frac{\partial F}{\partial M_{it}} = \frac{P_t^m}{P_t^y} \tag{6}$$

The first order necessary conditions with respect to capital and capital investment result in a simple Euler equation:

---

complicated functions and involve a greater burden of computation. This task is left for future work.

$$P_t^K \frac{\partial C^K}{\partial I_{it}} = \frac{1}{1+\rho} \mathrm{E}\left[P_{t+1}^Y \frac{\partial F}{\partial K_{it+1}} - P_{t+1}^K \left(1 + \frac{\partial C^K}{\partial K_{it+1}} \right) \right] \tag{7}$$

Similarly, the Euler equation for labor input and employment is:

$$P_t^L \frac{\partial C^L}{\partial E_{it}} = \frac{1}{1+\rho} \mathrm{E}\left[P_{t+1}^Y \frac{\partial F}{\partial L_{it+1}} - P_{t+1}^L \left(1 + \frac{\partial C^L}{\partial L_{it+1}} \right) \right] \tag{8}$$

On the other hand, the equation for the productivity level and R&D expenditure takes a slightly different form:

$$P_t^R \frac{\partial C^R}{\partial R_{it}} = \frac{1}{1+\rho} \mathrm{E}\left[\left(P_{t+1}^Y \frac{\partial F}{\partial \omega_{it+1}} - P_{t+1}^R \frac{\partial C^R}{\partial \omega_{it+1}} \right) \frac{\partial G}{\partial R_{it}} \right] \tag{9}$$

As can be seen, the decisions regarding inputs, M_{it}, and investment, I_{it}, E_{it}, and R_{it}, are functions of the state variables K_{it}, L_{it}, and ω_{it}. Moreover, unless production function F takes a special form, the decision rules are also functions of variable input and exogenous variables.

3. Estimating the production function

The discussion now turns to the more concrete step of the estimation of the production function. The production function in the previous section took an abstract form. Here, following the tradition of the literature on productivity, a Cobb-Douglas production function is assumed, which, moreover, makes it possible to compare the results obtained in this paper with those in other studies. The production function is assumed to take the following form:

$$Y_{it} = A_t K_{it}^{\beta_k} L_{it}^{\beta_l} M_{it}^{\beta_m} e^{\omega_{it}} e^{\varepsilon_{it}} \tag{10}$$

$$y = \beta_0 + \beta_{lt} l_{it} + \beta_{kt} k_{it} + \beta_{mt} m_{it} + \omega_{it} + \varepsilon_{it} \tag{11}$$

where the lower case indicates the logarithm of the corresponding variable. This is a very simple and standard functional form. If $\omega_{it} = \omega_i$, this is just a simple fixed effect model with a stochastic element. If ω_{it} is constant over time and the error term, ε_{it}, has autoregressive structure, it is a standard form of a Blundell and Bond (2000) regression. Their GMM regression (the so-called "system-GMM") is thought to be a solution to this kind of problem. However, this solution requires the assumption that productivity is constant over time.

Here, it is assumed that productivity is governed by a controlled first-order Markov process with transition probabilities $P(\omega_{it} | \omega_{it-1}, r_{it-1})$. This assumption is described in the following equation,

$$\omega_{it} = E[\omega_{it} \mid Info_{it-1}] + \xi_{it} = E[\omega_{it} \mid \omega_{it-1}, r_{it-1}] + \xi_{it} = g(\omega_{it-1}, r_{it-1}) + \xi_{it} \qquad (12)$$

where $Info_{it-1}$ is the information set in period t-1. This means that realized productivity of firm i in period t can be decomposed into two parts, that is, productivity expected from the information set at t -1, $g(\omega_{it-1}, r_{it-1})$ and a random shock, ξ_{it}, so that ξ_{it} is independent of r_{it-1}.[6]

As is well known, if the production function contains determinants (ω_{it}) which are not observed by the econometrician but observed by the manager of the firm, and if the observed inputs are chosen as a function of these determinants, then an endogeneity problem is present and OLS estimates of the coefficients on the observed inputs are biased.

To control for this, Olley and Pakes (1996) used investment as a proxy for productivity shocks. More recently, Levinsohn and Petrin (2003) have suggested using intermediate input as a proxy for productivity shocks instead. One of the reasons that they prefer intermediate input as a proxy is that intermediate input suffers less from the invertibility problem which is caused by the possible lumpiness of investment or zero investment. This paper follows the methodology adopted by Levinsohn and Petrin, that is, intermediate input is employed as a proxy.

However, the approach here differs from Levinsohn and Petrin's methodology with respect to the assumption of the productivity process. Their study assumes that productivity ω_{it} follows an exogenous Markov process. Typically, however, firms strive to enhance their productivity, often through R&D activities. To take this aspect into account, the productivity level is assumed to evolve according to the controlled first-order Markov process described in (12).

This assumption differs from those generally adopted in the literature in several respects. First of all, investment in R&D is assumed to raise productivity only in the next period. In the literature on knowledge capital, the investment in R&D in this period has a direct effect on the production in future periods. In this paper, investment in R&D in this period (r_{it}) has a direct effect on the production in the next period (ω_{it+1}), but not on the production in the following period or thereafter

[6] A random shock to productivity may be considered as the realization of the uncertainty related with productivity itself plus the uncertainties inherent in the R&D activity. In this sense, the shock is expected to be correlated with R&D expenditure in period $t-1$ in its variance. Discussion in this paper requires only the mean independency.

(ω_{it+2}, ω_{it+3}, ...). Although the investment in R&D in this period (r_{it}) contributes to productivity in period t+2 through the enhanced productivity in t+1,[7] the effect of the investment in R&D in t (r_{it}) on productivity in t+2 (ω_{it+2}) is not a direct one and cannot be identified. That is, this effect of r_{it} on ω_{it+2} is not distinguished in a qualitative way from the behavior of the productivity process itself, that is, the effect of ω_{it+1} on ω_{it+2}. However, this may not reflect the reality. Typically, firms engage in R&D employing a long time horizon, targeting profits three or five years ahead. To capture such a longer-term effect beyond period t+1, second or higher-order Markov processes should be introduced. Accommodating this idea is not impossible, but requires much more computation, so that this task is left for future work.

Second, the contribution of investment in R&D to productivity follows a stochastic process. As is well known, in the knowledge capital model, R&D expenditure is assumed to be accumulated in a deterministic way to become R&D stock. There is no uncertainty in this process.

Third, there is no need to construct R&D stock. In general, to construct reliable data on stock values, long time series of flow data are essential. However, long time-series data for R&D are rarely available. Because of the lack of data, many studies in the literature on knowledge capital therefore assume that the growth rate of R&D flow is equal to that of R&D stock. In addition, in the general knowledge capital model, the depreciation rate of knowledge is simply assumed to be a certain fixed level, such as 15 percent a year. But as the model in this paper uses only data on R&D expenditure, not R&D stock, there is no need for such assumptions.

The discussion now turns to the estimation process. The first step for the estimation is to choose which variable to use to proxy the unobserved productivity shock. As described above, in the estimation process of Levinsohn and Petrin (2003), intermediate inputs are used to proxy the productivity shock. Levinsohn and Petrin showed that under certain conditions, a firm's intermediate input has a monotone relationship with the firm's productivity level, just as Olley and Pakes (1996) proved the monotone relationship of investment in capital with the productivity level. Once the proxy variable is shown to be a monotone function of the productivity level, one can invert this function to express a firm's productivity level as a function of the capital stock and the proxy variable, whether the proxy variable is intermediate input or capital investment.

Since in Levinsohn and Petrin (2003) and Olley and Pakes (1996) labor input is a variable input, and is not used as a proxy, it is thought that the estimate of the coefficient of labor is not biased. For this reason, the first step in both papers is the estimation of the coefficient on labor by OLS estimation.

[7] In other words, the investment in R&D in this period (r_{it}) enhances the enhanced productivity in the next period (ω_{it+1}), and the productivity in the next period enhances the productivity in the period after next (ω_{it+2}).

But as Ackerberg, Caves, and Frazer (2006) point out, there is a possible collinearity problem. If labor input is a function of intermediate input, or if labor input is a fixed variable such as in the model in this paper, labor input shows up in the inverted productivity function. If this is the case, the coefficient on labor input is not identified because labor input shows up both in the inverted function and outside of the function. This is called the collinearity problem in this paper.

To make clear this problem of the unobservable productivity shock and input choice, consider the optimization condition of the Cobb-Douglas production function of equation (10).

The first-order necessary condition with respect to intermediate input with $E(\varepsilon_{it})=0$ gives the demand for intermediate input. By inverting this demand function, the productivity of the firm can be written as a function of capital, labor, intermediate input, and price of intermediate input:

$$\begin{aligned}\omega_{it} &\equiv h(k_{it}, l_{it}, m_{it}, p_t^m) \\ &= -\beta_0 - \ln\beta_m + (1-\beta_m)m_{it} - \beta_l l_{it} - \beta_k k_{it} + p_t^m.\end{aligned} \tag{13}$$

where $p_t^m = \ln P_t^m - \ln P_t$.

The price indexes do not have identification subscripts because perfect competition in factor markets is assumed. Using equation (13), production function (11) can be rewritten as[8]

$$y = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + h(k_{it}, l_{it}, m_{it}, p_t^m) + \varepsilon_{it} \tag{14}$$

Here, it is assumed that material is the only variable input. If labor is also assumed to be a variable input, as in most of the literature, equation (13) can be rewritten as follows using a first-order necessary condition with respect to labor:

$$\omega_{it} = \lambda_0 + (1-\beta_l-\beta_m)m_{it} - \beta_k k_{it} + (1-\beta_l)p_t^m + \beta_l p_t^l, \tag{15}$$

where $p_t^w = \ln P_t^w - \ln P_t$,

$p_t^m = \ln P_t^m - \ln P_t$, and

[8] Note again that both functions g in (12) and h in (14) do not have subscript for time t, which means that both functions are assumed to be time-invariant.

$$\lambda_0 = -\beta_0 - \ln\beta_m - \beta_l(\ln\beta_l - \ln\beta_m).$$

In this case, equation (11) can be rewritten as

$$y = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + h(k_{it}, m_{it}, p_t^m, p_t^l) + \varepsilon_{it} \tag{16}$$

If the inverted factor demand function (13) or (15) is inserted into equation (11), the production function is transformed into a simple optimization condition with the input terms being cancelled out. This equation means that the optimization conditions in a parametric estimation do not contribute to identification in this setting.

Another possibility is to estimate equation (14) or (16) by assuming function h in equation (14) or (16) as an unknown function. Fortunately, equation (16) can be estimated to obtain an estimate of the coefficient on labor. But as can be easily seen, estimation of (14) does not identify any coefficient of input because all of the inputs show up both in function h and outside of it. Except for such special cases, the production function cannot be estimated with inverted factor demand function h.

As mentioned above, in this step, Olley and Pakes (1996) and Levinsohn and Petrin (2003) estimate the coefficient on variable input because in their model, function h does not include labor, l. But as shown here, in most cases estimating the coefficient on labor in the first stage is problematic. In this paper, the coefficient on labor is not estimated in the first stage but in the second stage, as suggested by Ackerberg, Caves, and Frazer (2006).[9]

The purpose of this stage is to cancel out the random disturbance to obtain:

$$\Phi_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + h(k_{it}, l_{it}, m_{it}, p_t^m). \tag{17}$$

In this estimation, function h is thought to be an unknown function, because, as described above, optimization conditions do not contribute to the identification of the coefficients. One of the general ways to approximate an unknown function is to use a series estimator made up of a complete set of polynomials.[10] A series estimator of polynomials of degree three is used in this estimation.[11]

[9] In this paper, labor is assumed as a fixed input in the sense that it is decided at least before the current period begins. Later, in the estimation, this assumption of the fixity of labor is relaxed.

[10] When one uses an unknown function $h(u, v)$ of two variables u and v, a complete set of polynomials of degree x means a set of all polynomials of the form $u^a v^b$, where a and b are nonnegative integers such that $a+b \le x$. See Judd (1998).

To recover the implied ξ_{it}'s for any candidate value of the parameters $(\tilde{\beta}_k, \tilde{\beta}_l, \tilde{\beta}_m)$, the implied productivity shock $\omega_{it}(\tilde{\beta}_k, \tilde{\beta}_l, \tilde{\beta}_m)$ is computed as follows:

$$\tilde{\omega}_{it} = \hat{\Phi}_{it} - \tilde{\beta}_l l_{it} - \tilde{\beta}_k k_{it} - \tilde{\beta}_m m_{it} \tag{18}$$

Under the assumption of a controlled first-order Markov process in equation (12), regressing $\tilde{\omega}_{it}$ on a series estimator of polynomials of $\tilde{\omega}_{it-1}$ and r_{it-1} provides the implied ξ_{it}'s.

$$\tilde{\xi}_{it} = \tilde{\omega}_{it} - g(\tilde{\omega}_{it-1}, r_{it-1}) \tag{19}$$

The assumptions regarding the timing of the input decision yield the moment condition,

$$E\left[\xi_{it} + \varepsilon_{it} \middle| \begin{matrix} k_{it} \\ l_{it} \\ m_{it-1} \end{matrix}\right] = E\left[v_{it} \middle| \begin{matrix} k_{it} \\ l_{it} \\ m_{it-1} \end{matrix}\right] = 0 \tag{20}$$

where $v_{it} = \xi_{it} + \varepsilon_{it}$. Note that again, the capital stock and labor input are decided at time t -1 so that they do not respond to the innovation in productivity, and that last period's intermediate input decision should be uncorrelated with the innovation in productivity in the current period.

The sample analogue to the above moment condition is

$$\frac{1}{T}\frac{1}{N}\sum_t \sum_i \begin{pmatrix} k_{it} \\ l_{it} \\ m_{it-1} \end{pmatrix} \cdot v_{it}(\beta) = \frac{1}{T}\frac{1}{N}\sum_t \sum_i z_{it} \cdot v_{it}(\beta) \tag{21}$$

where $\beta = (\beta_k, \beta_l, \beta_m)'$.

[11] Previous studies show that polynomials of higher degree than four usually provide little more information on the original function.

Then, as described in Wooldridge (2002), the objective function can be written as follows:

$$\min_{\beta}\left[\frac{1}{\sqrt{N}}z'v(\beta)\right]' W\left[\frac{1}{\sqrt{N}}z'v(\beta)\right], \qquad (22)$$

where W is the weighting matrix, z is the matrix of z_{it}, $v(\beta)$ is the matrix of $v_{it}(\beta)$, and N is the number of firms during the period. The weighting matrix used in the first step for a consistent estimator is $W=\left(\frac{1}{N}z'z\right)^{-1}$. The second step uses the weighting matrix

$$W=\left(\frac{1}{N}z'v(\hat{\hat{\beta}})v(\hat{\hat{\beta}})'z\right)^{-1} \qquad (23)$$

where $\hat{\hat{\beta}}$ is the coefficient vector acquired in the first step. Function (23) is also used for the overidentification criteria.

4. Data

The main data set used in this paper is the data set of financial reports of listed firms compiled by the Development Bank of Japan (the "DBJ data set"). Most deflators used here for converting current values to real values are taken from the Japan Industrial Productivity Database 2006 (JIP 2006). A detailed description of the data construction is provided in the appendix.

The data used in this paper cover the period from 1999 to 2005. Even though most input and output data sets of firms are available since 1970, early R&D data appear highly unreliable, thus confining this study to a much shorter period. Only the accounting rules for R&D expenditure were changed in 1998, did data on R&D become much more reliable in Japan.

This study does not cover the entire economy because most R&D activity is concentrated in manufacturing industry. According to the *White Paper on Science and Technology*, until 2000, R&D expenditure in service industries was less than 10 percent of the total and even in 2006, R&D expenditure in service industries was no more than 12.6 percent of total R&D expenditure in Japan.

5. Estimation Results for the Production Function

5.1. Estimation equation

Given the set of assumptions spelled out above, the production function to be estimated is as follows:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + g(h_{it-1}, r_{it-1}) + \xi_{it} + \varepsilon_{it} \tag{24}$$

where

$$h_{it} = y_{it} - \beta_{lt} l_{it} - \beta_{kt} k_{it} - \beta_{mt} m_{it} \tag{25}$$

In this non-parametric estimation, function g is unknown. In this paper, as described above, a series estimator made up of a complete set of polynomials of degree three is used to estimate function g.

Another point to be noted here for the estimation of function g is that some firms do not perform R&D. A simple and common solution to take this into account is to use the following functional form:

$$g(h_{it-1}, r_{it-1}) = 1(r_{it-1} = 0) g_0(h_{it-1}) + 1(r_{it-1} > 0) g_1(h_{it-1}, r_{it-1}) \tag{26}$$

where

$$g_0(h_{it-1}) = g_{00} + g_{01}(h_{it-1}), \tag{27}$$

$$g_1(h_{it-1}, r_{it-1}) = g_{10} + g_{11}(h_{it-1}, r_{it-1}), \tag{28}$$

and 1(.) is the indicator function with the condition in the parenthesis. As described above, functions g_{01} and g_{11} are estimated with polynomials of up to degree three. One can easily see that the constants β_0, g_{01} and g_{11} are not estimated separately. Thus, in the estimation process, not β_0 is estimated but $\beta_0 + g_{01}$ and $\beta_0 + g_{11}$.

5.2. Instrumental Variables

According to equation (20), available instruments are current fixed inputs, K and L, one-period lagged variable input, M, and the lagged values of K, L, and M.[12] These values are used as instrument variables. Overidentification is tested with the criterion function (23).

5.3. Estimation Results

To compare the results with those in the literature, several traditional production functions are also estimated. The production functions estimated for comparison basically take the following functional form:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \kappa(n_{it}) + u_{it} \tag{29}$$

$$\kappa(n_{it}) = \gamma_0 1(n_{it} = 0) + 1(n_{it} > 0)(\gamma_1 + \gamma_n n_{it}) \tag{30}$$

where κ is a function related to knowledge input and n denotes the R&D stock. This functional form with the R&D stock follows the traditional knowledge capital model. R&D stock data are constructed as the accumulated value of R&D expenditure with a depreciation rate of 15 percent. To take non-R&D-performing firms into account, two dummy variables are added, one for R&D-performing and the other for non-R&D-performing firms.

Depending on the assumptions regarding the error term, u_{it}, the most commonly used estimation methods are OLS and fixed effect estimation (FXE henceforth). Here, both estimations were conducted, once with the function κ and once without it.[13]

Table 1 shows the results. Columns (1) and (3) do not include the term related to R&D. Columns (2) and (4) add the function κ related to R&D input to the production function elements.

Finally, for further comparison, column (5) shows the result of the estimation using Levinsohn and Petrin's approach (LP). Their estimation model does not include R&D as an input. As described above, their estimation scheme is well known as a way to control for endogeneity in the production function using material input as a proxy for productivity shocks.

---

[12] The products of the instruments mentioned above and the products of polynomials of the instruments are also available. How much additional information is available from these instrument variable series should be tested, but this task is left for future work. In this paper, the simplest tools are adopted.

[13] Random effect estimation was not performed here because the purpose is to compare coefficients. What matters is consistency, not efficiency.

Looking at the results in Table 1, the most notable point is that the estimates of the coefficient on capital are more significant in the estimation approach developed in this paper (labeled ENDOG) than in the alternative approaches. In many cases, using FXE, the result for the coefficient on capital is not significant or significant but negative, and in some instances, this is also the case using OLS. In contrast, using ENDOG, the coefficient is positive and significant in many cases.

These results require some elaboration. In most case, LP estimation did not obtain positive and significant estimates for the coefficient on capital. This may be interpreted as indicating that the differences of the estimates for the coefficient on capital arose from the differences in the two approaches, ENDOG and LP. As mentioned above, the differences are whether R&D is included and how to estimate the coefficient on labor.

To determine the origin of the difference in the coefficient estimates, two different estimations are conducted. The estimation labeled ENDOG1 (column (6) in Table 1) assumes that labor is a variable input, whereas ENDOG2 assumes that it is a fixed input. As the coefficient values show, the two results are almost identical. This demonstrates that the difference in the estimation results does not seem to be attributable to the fixity of labor. The remaining possible origins of this difference therefore are the inclusion of R&D or the collinearity problem described in Section 3.

Comparing the results of the estimation using the method of LP and ENDOG shows that both factors are responsible for the difference. A possible reason is that in the LP estimation, the labor coefficient is estimated in the first stage, so that the coefficient value has nothing to do with R&D input. If the difference is mostly attributable to the R&D input, then the coefficient value of labor in ENDOG should be identical or similar to that of the LP estimation. But as seen in Table 1, the labor coefficient values of the ENDOG estimation are quite different from that of the LP estimation. Taking into account what Ackerberg, Caves, and Frazer (2006) point out, both collinearity and the absence of R&D input are likely to be the main causes of the difference in the capital coefficient estimates.

A final issue to be considered here are data problems related to R&D. The main data set used in this paper contains five items related to R&D expenditure. Two items concern the cost expended for research activities, two refer to the cost of development activities, and the fifth item is the aggregated value. Accounting rules regarding R&D expenditure changed in 1998. Before that year, it was not compulsory to report R&D expenditure and, consequently, only some firms reported it.

To mitigate this problem, the production function was estimated again using only data of firms that either reported R&D expenditure in all years or that reported no R&D expenditure in all years throughout the observation period (i.e., firms that report R&D expenditure in some years, but not in others, were excluded). The results, employing again the LP estimation and the two ENDOG estimations – with labor treated as a variable input in the first and as a fixed input in the second – are shown in Table 2. The results are essentially the same as the main results reported in Table 1.

6. Productivity and R&D Investment

6.1. Productivity comparison

This section examines the characteristics of the productivity index calculated using the results of the main estimation and the relationship of the index with R&D investment. In the context of this paper, the productivity of each firm ($\hat{\omega}_{it}$) is defined as follows:[14]

$$\hat{\omega}_{it} = y_{it} - \hat{\beta}_{lt} l_{it} - \hat{\beta}_{kt} k_{it} - \hat{\beta}_{mt} m_{it} \quad (31)$$

For the comparison, two indexes of total factor productivity are calculated for each firm every year. The first of these indexes is calculated employing the methodology developed by Good, Nadiri, and Sickles (1997) and Aw, Chen, and Roberts (2001). This index basically measures the distance of a firm's productivity from the average productivity of the industry in the *base year* (1981 in this paper),[15] and is labeled as *lnTFP1* in this paper. The second index measures the distance of a firm's productivity from the industry average in the *current year* and is denoted as *lnTFP2* hereafter.[16]

The summary statistics of the TFP indexes *w*, *lnTFP1*, and *lnTFP2* are shown in Table 3, while the correlation between them is presented in Table 4. The latter indicates that the main productivity index is significantly correlated with those used for comparison.

The purpose here is to examine the origin of the difference between the indices – if any significant differences exist – focusing on the relationship between *lnTFP1* and ω. To this end, it is assumed for the time being that ω is the true index for total factor productivity and the biasedness of a productivity index is defined by comparing the index with ω. A simple way to do this is as follows. After regressing *lnTFP1* on ω, the predicted value for each ω is calculated. The case where the

---

[14] Strictly speaking, the right-hand side of (30) corresponds to $\hat{\omega}_{it} + \hat{\varepsilon}_{it}$. However, for notational simplicity it is simply expressed as $\hat{\omega}_{it}$. ω without the subscripts is used when no ambiguities arise as a result.

[15] Data used for the estimation covers from 1999 to 2005. But the limitation is because of the credibility of R&D data. Other data has much longer coverage. TFP is calculated with data from 1981 to 2005.

[16] A more detailed explanation of the indexes of TFP is found in the appendix.

actual *lnTFP* is greater than the predicted one is defined as an upward bias and the opposite as a downward bias.

Table 5 reports the results. Columns (1) and (2) show the means of the productivity ω of the two groups, the upwardly-biased and the downwardly-biased group.

The two groups are expected to have the same mean and Table 5 shows this. Looking at the results for each industry, *t*-tests for comparing the means of the productivity of the two groups indicate that there are no significant differences in the productivity levels in five of the seven industries. The two industries in which the differences are significant are the wholesale and retail industry and the chemical industry, but the actual magnitudes of the differences are rather small.

Comparing the ways productivity of ω and *lnTFP1* are measured, the biggest difference is whether cost shares of each input are estimated or calculated. This issue is examined more closely in columns (3) to (14).

The cost shares in columns (3), (7), and (11) which are used to calculate *lnTFP1* differ in at least three respects from the coefficient estimates, β's, in columns (4), (8), and (12) which are used to calculate ω. The first is that the cost shares are calculated based on how much is expended for the input factors, so that the cost shares are not immune to the endogeity problem, whereas the coefficient estimates are expected to be. The second is that the cost share is based on constant returns to scale so that the cost shares sum up to 1. The third is that cost shares may differ by firm because they use individual firm's cost structure of labor, capital, and intermediate input, whereas coefficients of the inputs are estimated by industry, by assuming that the coefficients on each input are the same fall all firms in an industry.

In general, one can see that the cost share of labor in column (3) is overestimated, whereas the cost share of capital in column (7) seems to be underestimated. In the case of intermediate input, it is not clear which is the case.

If the cost-based share is overestimated, what kind of bias is caused? Columns (5) and (6) compare the averages of the cost share of labor between the downwardly-biased and the upwardly-biased group. In most cases, the downwardly-biased group has a higher cost share of labor, meaning that the *lnTFP1* underestimates the productivity of more labor-intensive firms.

As for the cost share of capital, comparing columns (9) and (10) does not reveal as clear a relationship between the cost share of capital and the direction of bias as in the case of cost share of labor. Although *lnTFP1* underestimates the cost share of capital, the effect of the bias differs depending on the industry. In the general machinery, electrical and electronic machinery, and transportation machinery industries, the productivity of more capital intensive firms tends to be downwardly biased, whereas in the other industries, the productivity of capital intensive firms tends to be downwardly biased.

6.2. R&D investment and productivity

From the definition of the production function in (11), the simplest way to define the returns to R&D investment is as follows:[17]

$$\frac{\partial y_t}{\partial r_{it-1}} = \frac{\partial g_1(h_{it-1}, r_{it-1})}{\partial r_{it-1}} = \frac{\partial g_{11}(h_{it-1}, r_{it-1})}{\partial r_{it-1}} \tag{32}$$

As described above, the values of h_{it-1} and r_{it-1} differ by firm and year, so that the value in (32) also differs by firm and year. Table 6 shows the distribution of R&D returns by industry. Returns vary widely not only from industry to industry, but even within an industry. This means that some firms enjoy higher R&D returns than others.

This raises the question: what determines the returns on R&D investment? To analyze this, R&D returns are regressed on firm characteristics, that is, firms' size, return on assets (ROA), debt ratio, and ownership structure. Table 7 shows the results. In four industries of the seven, R&D returns turn out to be correlated with firms' size as measured by the volume of sales. In other words, the bigger a firm is, the greater are the benefits from R&D activity. Figure 2 shows these relationship between R&D returns and firms' size by industry.[18]

Another point to note is that R&D returns are positively correlated with ROA. In five of the seven industries, this correlation is clear and strong, meaning that more profitable firms enjoy higher returns from their R&D investment.

R&D returns seem to be negatively correlated with the debt ratio and the ratio of shares owned by the government (labeled GOV in the table 7). On the other hand, R&D returns are positively correlated with the ratio of shares owned by foreign firms (labeled FRN in the table 7) and the ratio of shares owned by private investors (labeled PER in the table 7), even though those relationship is weak.

Another point of interest is how R&D returns are viewed by the stock market. Table 8 shows that, in general, Tobin's q is positively correlated with R&D returns.[19]

Return on assets (ROA) is included as a control variable. ROA is thought of as a main factor

[17] This definition is different from that in the knowledge capital literature. Estimation of the returns to R&D in the knowledge capital model captures the rate of return to R&D stock, not R&D flows, whereas in this paper, the main estimation captures the returns to R&D flow. R&D stock is not defined here.

[18] In the chemical industry, two distinct clusters can be observed. A possible reason is that this industry includes heterogeneous sub-industries, such as pharmaceuticals and non-pharmaceuticals. However, even when plotting the charts for pharmaceuticals and non-pharmaceuticals separately, as shown in the figure, the same clustering is observed.

[19] The definition and method of constructing Tobin's q are described in detail in the appendix.

affecting the share price. This result shows that even when controlling for the effect of ROA, R&D returns tend to significantly affect the share price. Variables representing firm size, such as sales and the number of employees, are not included in this regression because when they were included, their coefficients were generally insignificant.

7. Concluding remarks

This paper attempted to apply recently developed econometric methods to control for the endogeneity problems and estimation biases arising from missing variables. The idiosyncratic productivity shock that causes the endogeneity problem between the inputs and output is replaced with an inverted intermediate input demand function as in Levinsohn and Petrin (2002). But as Ackerberg, Caves, and Frazer (2006) suggested, the inverted demand function is thought to be an unknown function and the coefficient on labor is estimated in the second stage. Firms are assumed to conduct R&D to enhance the productivity of the next period, and the productivity process follows a controlled first-order Markov process.

This paper found that there are estimation biases and that the possible origins of the biases are unobservable productivity shocks (endogeneity problem) and ignoring the contribution of R&D activity. The biases are especially prominent in the estimates of the coefficient on capital.

Calculating the returns on R&D investment using the estimation results, it was found that R&D returns are positively correlated with firm size (measured by sales) and ROA (return on assets), and that markets take into account this R&D return in their valuation of firms.

Appendix

This appendix provides a detailed description of how the data set was constructed.

Output

For output, sales after adjusting for inventory are used. For the wholesale and retail industry, purchases of merchandise are subtracted from sales. The price index for output and input is taken from the JIP2006 data base.[20]

Price of Capital Goods

Capital goods consist of the following six types of assets:

(1) nonresidential buildings;

(2) structures;

(3) machinery;

(4) transportation equipment;

(5) instruments and tools; and

(6) land.

The price index used for deflating (1) and (2) is that for construction materials in the corporate goods price index (CGPI). For machinery, the weighted average of the following three CGPI components was used: general machinery and equipment, electrical machinery and equipment, and precision instruments. As the (fixed) weight, the capital formation matrices for 1985, 1990, 1995, and 2000 rearranged by the Research Institute of Economy, Trade and Industry (RIETI) by industry are used. The same procedure is employed to construct the price index for instruments and tools. The price index for instruments and tools is the weighted average of five CGPI components: metal products, general machinery and equipment, electrical machinery and equipment, precision instruments, and other manufacturing industry products. Again, the capital formation matrices are used as the fixed weight. The transportation equipment component of the CGPI is adopted as the price index for transportation equipment. Finally, for land, the index of urban land prices compiled by the Japan Real Estate Research Institute is used. The index for commercial areas is adopted for non-manufacturing firms, whereas that for industrial areas is adopted for manufacturing firms.

Nominal investment

The following notations are used in the calculation of nominal investment:

KGB_t: book value of gross capital stock at the end of the period;

KNB_t: book value of net capital stock at the end of the period;

[20] The JIP2006 data base provides deflators up to 2002. They were extended here up to 2004 using SNA deflators.

AD_t: book value of accumulated depreciation ;

DEP_t: accounting depreciation during the period.

The definition of nominal investment is:

$$NOMI_t = KNB_t - KNB_{t-1} + DEP_t \quad \text{(A.1)}$$

Since DEP_t is not available until 1977,[21] (AD_t - AD_{t-1}) was used as a weight to distribute total depreciation between the five kinds of capital goods excluding land.

Capital stock

The perpetual inventory method is used to calculate real capital stock:

$$K_t = (1-\delta)K_{t-1} + \frac{NOMI_t}{PK_t} \quad \text{(A.2)}$$

where PK_t is the price index for the capital asset. The initial year chosen for the calculation based on the perpetual inventory method is 1970, because accumulated depreciation is only available since 1969. In the main regressions, capital stock does not include land. However, to construct Tobin's q, land stock is calculated. To convert the book value of land to the market value, a somewhat complicated procedure was adopted.

Using the book value of land stock at the end of each period and the acquisition of land during the year, it is possible to calculate the acquisition value of the land acquired during the period. However, the statistics do not allow to discern when the land sold during this period was acquired, so that it is not clear how to apply the price index for land to the land value sold during the period.

For this reason, the "last-in-first-out" principle is assumed for land. That is, when firms sell land, it is assumed that they sell the land which was acquired last. Accumulated net purchases are calculated backward and it is assumed that the land sold during this period was acquired during the period when the accumulated net purchase first turns positive.

Here is an example.

Year	Bought	Sold	acc1998	acc1999
1991	100		400	220
1992	120		300	120
1993	160	30	180	0

[21] Depreciation by asset is only available after 1978. Before 1977, the sum of the depreciation for all assets is reported.

1994		110	50	-130
1995	170	100	160	-20
1996	80		90	-90
1997	100		10	-170
1998		90	-90	-270
1999		180		-180

Variable "acc1998" is the accumulated value of land which is bought in the current period (in this example 1998) or has been bought in the past (in this example 1997, or earlier) from the current period to the past. Land which is sold is added to the sum as a negative value. Thus, this variable should be read from the current to the past. This variable, therefore, shows how many periods ago the land was bought which is sold in the current period. Variable acc1999 is defined in the same way.

The land sold in 1998 (90 units of land) was bought in 1997. When one looks at acc1998 and reads from 1998 backwards, it first turns positive in 1997. In 1999, 180 units of land were sold. Under the last-in-first-out principle, the land sold in 1999 includes the land which was bought in 1997, 1996, 1995 and 1993. In this case, the price index of land in 1993 is applied to the land which is sold in 1999.

Depreciation rate

The JIP2006 provides fixed capital formation matrixes aggregated to 39 assets by JIP2006 industry classification and corresponding depreciation rates. Aggregate depreciation rates for the five capital goods are calculated using the industry weights from the fixed capital formation matrix. The average depreciation rates are (1) 8.31297%, (2) 2.25949%, (3) 12.77375%, (4) 17.12287%, and (5) 12.45546%.[22]

Capital stock aggregation

A Divisia index or Tronqvist index should be applied here. However, once the base year is set, the productivity of firms which did not exist in that year cannot be calculated. For this reason, the capital stock is aggregated by summing up the market value of each type of capital good.

Capital cost

Capital cost is measured as follows:

$$c_k = \frac{1-z}{1-u} p_k \{\lambda r + (1-u)(1-\lambda)i + \delta - (\frac{\dot{p}_k}{p_k})\} \tag{A.3}$$

[22] For comparison, Hayashi and Inoue (1991) use depreciation rates of (1) 4.7%, (2) 5.64%, (3) 9.489%, (4) 14.70%, and (5) 8.838%.

where z is the expected present value of tax savings due to depreciation allowances on a yen of investment in capital goods, u is the efficient corporate tax rate, λ is the own-capital ratio (=1-debt/total asset), r is the long-term bond rate, i is the prime rate, δ is depreciation, and p_k is the price index.

z is calculated as follows:

$$z = (u \cdot \delta) / [\{\lambda r + (1-u)(1-\lambda)i\} + \delta]. \qquad \text{(A.4)}$$

Tax saving, z, is not calculated for land stock because land has no depreciation. Thus, capital cost for land, c_{land}, is slightly different from the one for other capital goods for the same reason.

Capital cost is calculated by multiplying c by the capital stock. Labor costs and material costs are obtained from profit/loss tables.

Effective corporate tax rate

Following Hayashi and Inoue (1991), the effective corporate tax is calculated as follows:

$$t = \frac{(u_t + v_t)(1 + r_t)}{(1 + r_t + v_t)} \qquad \text{(A.5)}$$

where u_t is the corporate tax rate, v_t is the enterprise tax rate, and r_t is the short-term interest rate.

Labor input

Man-hours are used here as labor input. Labor hour data are taken from the JIP2006 data base and extended up to 2004 using the *Monthly Labor Survey*. Industry average man-hours are applied to each firm classified in that industry because firm data for labor hours are not available.

Intermediate input

Intermediate input is calculated as follows:

Sales cost + Selling, general and administrative expenses
– Depreciation – Increase of product – Increase of goods in process

For the retail and wholesale sector, purchases of merchandise are subtracted from this intermediate input. The price index for intermediate input is taken from the JIP2006 data base.[23]

[23] The JIP2006 data base provides deflators up to 2002, which were extended up to 2004 using SNA deflators.

Calculating TFP

Following Good, Nadiri, and Sickles (1997) and Aw, Chen, and Roberts (2001), the TFP level of firm *f* in year *t* in a certain industry is calculated in comparison with the TFP level of a hypothetical representative firm in year 0 in that industry by

$$\ln TFP1_{f,t} = (\ln Q_{f,t} - \overline{\ln Q_t}) - \sum_{i=1}^{n} \frac{1}{2}(S_{i,f,t} + \overline{S_{i,t}})(\ln X_{i,f,t} - \overline{\ln X_{i,t}})$$
$$+ \sum_{s=1}^{t} (\overline{\ln Q_s} - \overline{\ln Q_{s-1}}) - \sum_{s=1}^{t} \sum_{i=1}^{n} \frac{1}{2}(\overline{S_{i,s}} + \overline{S_{i,s-1}})(\overline{\ln X_{i,s}} - \overline{\ln X_{i,s-1}})] \quad \text{(A.6)}$$

where $Q_{f,t}$, $S_{i,f,t}$, and $X_{i,f,t}$ denote the gross output of firm *f* in year *t*, the cost share of factor *i* for firm *f* in year *t*, and firm *f*'s input of factor *i* in year *t*, respectively. Variables with an upper bar denote the industry average of that variable.

Constant returns to scale are assumed. As factor inputs, capital, labor and real intermediate inputs are taken into account. The representative firm for each industry is defined as a hypothetical firm whose logarithmic value of gross output as well as the logarithmic value of inputs and cost shares of all production factors are identical with the industry averages.

The first two terms on the right-hand side of equation (A.6) denote the gap between firm *f*'s TFP level in year *t* and the representative firm's TFP level in that year. The third and fourth term denote the gap between the representative firm's TFP level in year *t* and the representative firm's TFP level in year *0*. Therefore, $lnTFP_{f,t}$ in equation (A.6) denotes the gap between firm *f*'s TFP level in year *t* and the representative firm's TFP level in year 0.

Cross-sectional TFP (*lnTFP2*) is defined in a simpler way:

$$\ln TFP2_{f,t} = (\ln Q_{f,t} - \overline{\ln Q_t}) - \sum_{i=1}^{n} \frac{1}{2}(S_{i,f,t} + \overline{S_{i,t}})(\ln X_{i,f,t} - \overline{\ln X_{i,t}}) \quad \text{(A.7)}$$

TFP is calculated using equations (A.6) and (A.7), with 1980 used as the base year. Observations whose deviation of *lnTFP1* from the industry average of *lnTFP1* in a year is greater than three times the industry standard deviation of *lnTFP1* in that year is thought to be outliers and are discarded. Then *lnTFP1* and *lnTFP2* are calculated again with re-calculated values of the industry averages of inputs, output, and cost shares.

Tobin's q

Tobin's q is defined as the sum of the market value of equity and debt divided by the replacement cost of capital. The market value of equity is calculated as the number of stocks issued multiplied by the stock price. The stock price is at the first transaction day of the month following the financial report. If no price information for that day is available, the price for the earliest date following the financial report is used. Debt includes only liabilities with interest. The replacement cost of capital is calculated as the sum of the market value of aggregated capital above and the total sum of assets minus the book value of tangible fixed assets.

References

Ackerberg, D., K. Caves, and G. Frazer, 2006. Structural identification of production functions. Working paper. Los Angeles: UCLA.

Arellano, M. and O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1): 29-51.

Aw B. Y., X. Chen, and M. J. Roberts. 2001. Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing. *Journal of Development Economics* 66: 55-86.

Blundell, R., and S. Bond. 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Review* 19: 321-340.

Buettner, T. 2005. Productivity Dynamics when Labor is a State Variable. Working paper. London: London School of Economics.

Doraszelski, U., and J. Jaumandreu. 2007. R&D and productivity: Estimating production functions when productivity is endogenous. Working paper. Cambridge, MA: Harvard University.

Good D. H., Nadiri M. I., and Sickles R. C. 1997. Index number and factor demand approaches to the estimation of productiviy. In *Handbook of Applied Econometrics: Microeconomics.* Vol. 2. ed. Pesaran, H. and P. Schmidt. Oxford: Blackwell.

Griliches, Z. 1981. Market Value, G&D, and Patents. *Economic Letters* 7: 183-87.

———. 1996. *R&D and Productivity: The Econometric Evidence*, Chicago: Chicago Univ. Press.

Hayashi, F., and T. Inoue. 1991. The Relation between Firm Growth and *Q* with Multiple Capital Goods: Theory and Evidence from Panel Data on Japanese Firms. *Econometrica* 59 (3): 731-53.

Judd, K. 1998. Numerical methods in economics. Cambridge: MIT Press.

Levinsohn, J., and A. Petrin. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies* 70: 317-41.

Marschak, J., and W. Andrews. 1944. Random simultaneous equations and the theory of production. *Econometrica* 12 (3): 143-205.

Ministry of Education, Culture, Sports, Science and Technology. 2006. *WHITE PAPER ON SCIENCE AND TECHNOLOGY 2006*. Tokyo: Ministry of Education, Culture, Sports, Science and Technology, Japanese Government.

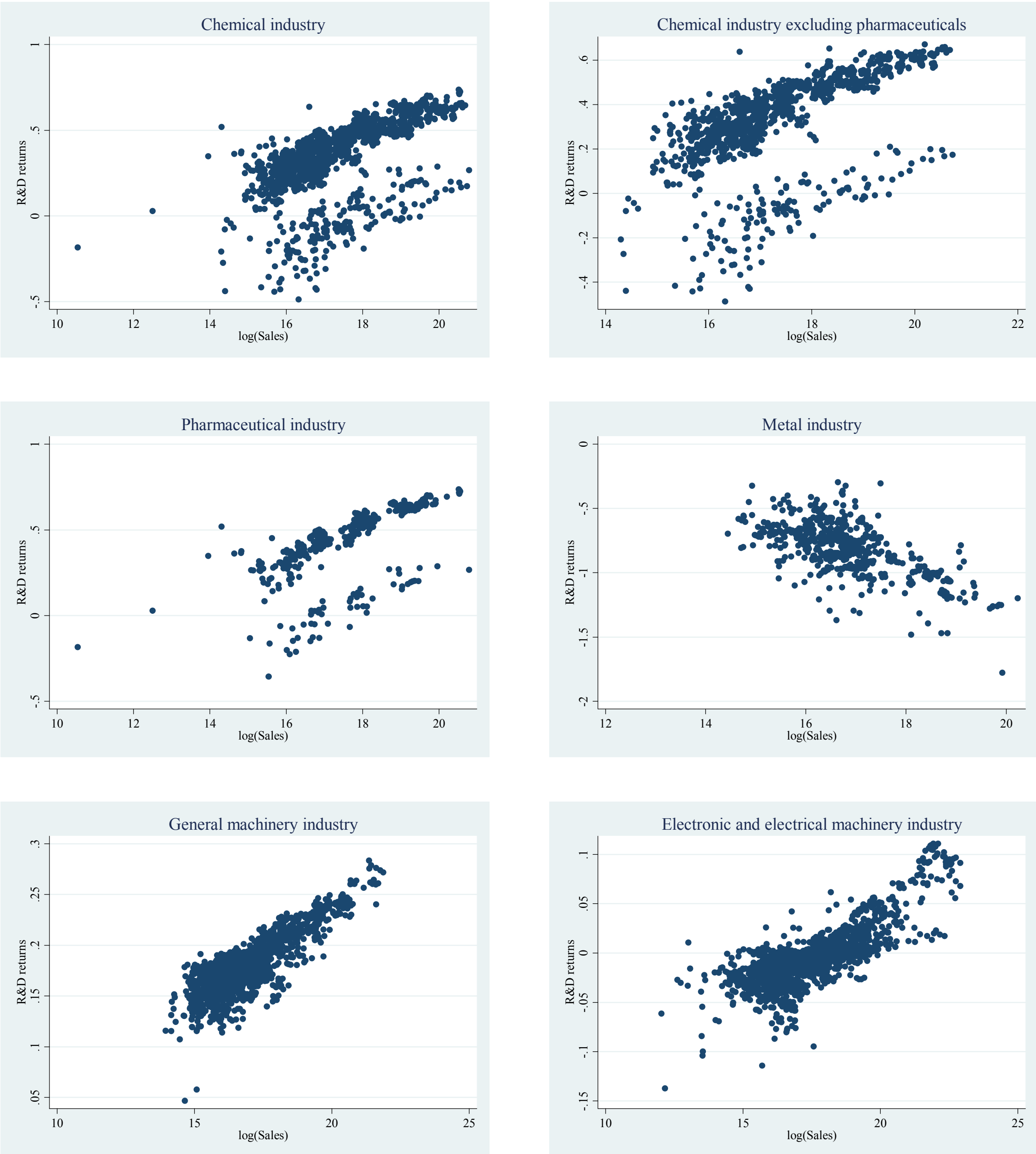
Olley, S. and A. Pakes. 1996. The dynamics of productivity in the telecommunications industry. *Econometrica* 64 (6): 1263-97.

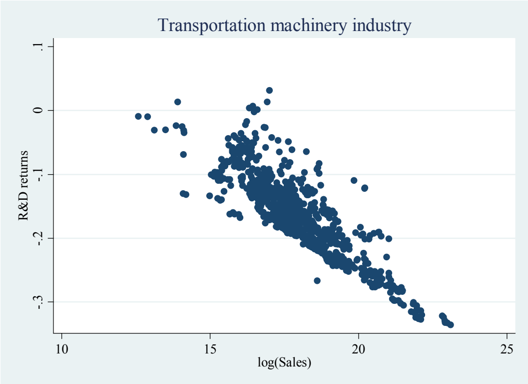
Wildasin, D. E. 1984. The *q* theory of Investment with Many Capital Goods. *American Economic Review* 74 (1): 203-10.

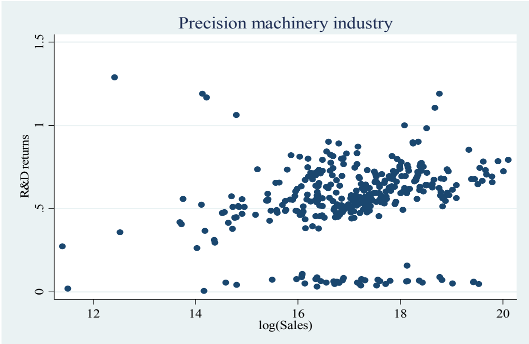
Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Figure 1. Comparison of Productivity Indexes

Figure 2. R&D Returns and Sales




Transportation machinery industry
R&D returns
.1
0
-.1
-.2
-.3
10
15
20
25
log(Sales)


Precision machinery industry
R&D returns
1.5
1
.5
0
12
14
16
18
20
log(Sales)

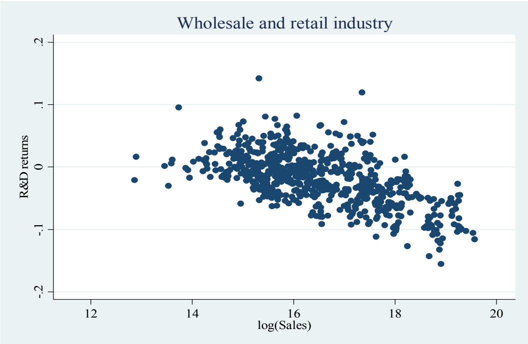

Wholesale and retail industry
R&D returns
.2
.1
0
-.1
-.2
12
14
16
18
20
log(Sales)

Table 1. Production function estimation

		OLS1	OLS2	FXE1	FXE2	LP	ENDOG1	ENDOG2
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Manufacturing industry	Labor	0.138 ***	0.123 ***	-0.032 ***	-0.030 ***	0.142 ***	0.043 *	0.174 ***
		(0.004)	(0.004)	(0.005)	(0.005)	(0.008)	(0.025)	(0.021)
	Capital stock	0.003	-0.002	0.003	0.000	0.010 ***	0.043 ***	0.043 ***
		(0.002)	(0.002)	(0.005)	(0.005)	(0.000)	(0.000)	(0.000)
	Material	0.858 ***	0.841 ***	0.857 ***	0.855 ***	0.330 **	0.913 ***	0.826 ***
		(0.003)	(0.003)	(0.004)	(0.004)	(0.145)	(0.039)	(0.018)
	Dummy (w/o R&D)		1.310 ***		(dropped)			
			(0.033)					
	Dummy (w/ R&D)		0.791 ***		-0.127 ***			
			(0.023)		(0.025)			
	R&D stock		0.037 ***		0.009 ***			
			(0.002)		(0.002)			
	Constant	0.766 ***		3.160 ***	3.180 ***			
		(0.023)		(0.084)	(0.084)			
	R-squared	0.978	1.000	0.859	0.859			
	No. of observations	11724	11724	11724	11724	11724	11724	11724
Chemical industry	Labor	0.207 ***	0.202 ***	-0.046 ***	-0.043 ***	0.179 ***	0.087	0.043
		(0.011)	(0.013)	(0.017)	(0.017)	(0.025)	(0.122)	(0.056)
	Capital stock	0.093 ***	0.093 ***	-0.013	-0.019	0.010	0.348 ***	0.348 ***
		(0.007)	(0.007)	(0.017)	(0.017)	(0.024)	(0.011)	(0.088)
	Material	0.725 ***	0.725 ***	0.687 ***	0.686 ***	0.980 **	0.783 ***	0.391 ***
		(0.009)	(0.010)	(0.012)	(0.012)	(0.414)	(0.104)	(0.103)
	Dummy (w/o R&D)		0.696 ***		0.254 **			
			(0.109)		(0.100)			
	Dummy (w/ R&D)		0.634 ***		(dropped)			
			(0.061)					
	R&D stock		0.004		0.017 **			
			(0.005)		(0.007)			
	Constant	0.627 ***		6.540 ***	6.350 ***			
		(0.060)		(0.277)	(0.288)			
	R-squared	0.982	1.000	0.802	0.803			
	No. of observations	1533	1533	1533	1533	1533	1533	1533
Metal industry	Labor	0.169 ***	0.150 ***	0.155 ***	0.150 ***	0.107 ***	0.043	0.217 ***
		(0.011)	(0.011)	(0.019)	(0.019)	(0.025)	(0.106)	(0.029)
	Capital stock	0.001	-0.001	0.021	0.020	0.980 ***	0.043 **	0.043 ***
		(0.007)	(0.007)	(0.016)	(0.016)	(0.357)	(0.020)	(0.008)
	Material	0.788 ***	0.782 ***	0.780 ***	0.780 ***	0.980 ***	0.478 ***	0.913 ***
		(0.009)	(0.009)	(0.010)	(0.010)	(0.035)	(0.081)	(0.038)
	Dummy (w/o R&D)		1.860 ***		0.030			
			(0.089)		(0.061)			
	Dummy (w/ R&D)		1.430 ***		(dropped)			
			(0.073)					
	R&D stock		0.032 ***		0.001			
			(0.004)		(0.005)			
	Constant	1.480 ***		1.500 ***	1.560 ***			
		(0.074)		(0.250)	(0.261)			
	R-squared	0.984	1.000	0.930	0.931			
	No. of observations	702	702	702	702	702	702	702
General machinery industry	Labor	0.162 ***	0.155 ***	0.020 **	0.021 **	0.162 ***	0.174 ***	0.130 ***
		(0.008)	(0.008)	(0.009)	(0.009)	(0.028)	(0.029)	(0.047)
	Capital stock	0.020 ***	0.014 **	-0.021 ***	-0.024 ***	0.010	0.043 ***	0.043 **
		(0.005)	(0.005)	(0.008)	(0.008)	(0.123)	(0.000)	(0.018)
	Material	0.820 ***	0.806 ***	0.967 ***	0.967 ***	0.700 ***	0.783 ***	0.870 ***
		(0.005)	(0.005)	(0.009)	(0.009)	(0.110)	(0.038)	(0.033)
	Dummy (w/o R&D)		1.180 ***		0.192 ***			
			(0.064)		(0.045)			
	Dummy (w/ R&D)		0.832 ***		(dropped)			
			(0.039)					
	R&D stock		0.025 ***		0.014 ***			
			(0.003)		(0.003)			
	Constant	0.765 ***		0.928 ***	0.776 ***			
		(0.038)		(0.188)	(0.191)			
	R-squared	0.991	1.000	0.899	0.901			
	No. of observations	1689	1689	1689	1689	1689	1689	1689

1. Estimation period : 1999-2005.
2. FXE denotes Fixed Effect Estimation, LP denotes the estimation following Levinsohn and Petrin (2003), and ENDOG denotes the estimation methodology developed in this paper.
3. In (6) labor is a variable input, whereas in (7) it is a fixed input.
4. *, **, and *** indicate $p < 0.1$, $p < 0.5$, and $p < 0.01$, respectively.

Table 1. Production function estimation (cont.)

		OLS1	OLS2	FXE1	FXE2	LP	ENDOG	ENDOG
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Electronic and electrical machinery industry	Labor	0.155 ***	0.148 ***	-0.008	0.000	0.153 ***	0.130 ***	0.130 ***
		(0.016)	(0.016)	(0.019)	(0.019)	(0.037)	(0.024)	(0.036)
	Capital stock	-0.067 ***	-0.075 ***	-0.080 ***	-0.088 ***	0.010	0.087 ***	0.087 ***
		(0.010)	(0.010)	(0.019)	(0.019)	(0.134)	(0.000)	(0.000)
	Material	0.941 ***	0.915 ***	0.983 ***	0.976 ***	0.250	0.913 ***	0.913 ***
		(0.011)	(0.011)	(0.016)	(0.017)	(0.217)	(0.036)	(0.048)
	Dummy (w/o R&D)		0.951 ***		(dropped)			
			(0.137)					
	Dummy (w/ R&D)		0.259 ***		-0.357 ***			
			(0.084)		(0.113)			
	R&D stock		0.045 ***		0.024 ***			
			(0.007)		(0.007)			
	Constant term	0.274 ***		2.020 ***	2.170 ***			
		(0.084)		(0.357)	(0.359)			
	R-squared	0.958	1.000	0.726	0.728			
	No. of observations	1898	1898	1898	1898	1898	1898	1898
Transportation machinery industry	Labor	0.170 ***	0.166 ***	0.040 ***	0.051 ***	0.162 ***	0.043 *	0.130 ***
		(0.006)	(0.006)	(0.008)	(0.008)	(0.015)	(0.026)	(0,024)
	Capital stock	0.008 *	0.008 *	-0.016	-0.019 *	0.010	0.087 ***	0.087 ***
		(0.004)	(0.004)	(0.010)	(0.010)	(0.277)	(0;.013)	(0.008)
	Material	0.826 ***	0.813 ***	0.904 ***	0.892 ***	0.980 ***	0.913 ***	0.870 ***
		(0.005)	(0.005)	(0.006)	(0.007)	(0.177)	(0.008)	(0.025)
	Dummy (w/o R&D)		0.996 ***		0.116 ***			
			(0.045)		(0.023)			
	Dummy (w/ R&D)		0.786 ***		(dropped)			
			(0.027)					
	R&D stock		0.015 ***		0.009 ***			
			(0.002)		(0.001)			
	Constant	0.724 ***		1.650 ***	1.630 ***			
		(0.027)		(0.187)	(0.182)			
	R-squared	0.998	1.000	0.961	0.963			
	No. of observations	1006	1006	1006	1006	1006	1006	1006
Precision machinery industry	Labor	0.202 ***	0.203 ***	0.057 **	0.059 **	0.185 ***	0.174	0.130 ***
		(0.022)	(0.022)	(0.024)	(0.024)	(0.064)	(0.136)	(0.098)
	Capital stock	-0.016	-0.019	-0.032	-0.036	0.010	0.087 ***	0.348 ***
		(0.017)	(0.018)	(0.026)	(0.026)	(0.126)	(0.008)	(0.008)
	Material	0.822 ***	0.817 ***	1.060 ***	1.060 ***	0.140	0.913 ***	0.435 ***
		(0.016)	(0.017)	(0.023)	(0.024)	(0.344)	(0.161)	(0.076)
	Dummy (w/o R&D)		0.859 ***		(dropped)			
			(0.197)					
	Dummy (w/ R&D)		0.727 ***		-0.149			
			(0.115)		(0.228)			
	R&D stock		0.008		0.009			
			(0.011)		(0.015)			
	Constant	0.736 ***		-0.975 **	-0.904 *			
		(0.114)		(0.469)	(0.479)			
	R-squared	0.982	1.000	0.867	0.868			
	No. of observations	394	394	394	394	394	394	394
Wholesale and retail industry	Labor	0.233 ***	0.233 ***	0.106 ***	0.106 ***	0.247 ***	0.217 ***	0.261 ***
		(0.006)	(0.006)	(0.007)	(0.007)	(0.019)	(0.039)	(0.039)
	Capital stock	0.027 ***	0.027 ***	0.034 ***	0.034 ***	0.010	0.043 **	0.043
		(0.004)	(0.004)	(0.007)	(0.007)	(0.032)	(0.022)	(0.035)
	Material	0.695 ***	0.694 ***	0.665 ***	0.665 ***	0.320	0.739 ***	0.696 ***
		(0.005)	(0.005)	(0.007)	(0.007)	(0.234)	(0.045)	(0.043)
	Dummy (w/o R&D)		1.960 ***		0.072			
			(0.047)		(0.087)			
	Dummy (w/ R&D)		1.800 ***		(dropped)			
			(0.082)					
	R&D stock		0.013 **		0.004			
			(0.006)		(0.007)			
	Constant	1.930 ***		4.020 ***	3.950 ***			
		(0.046)		(0.107)	(0.135)			
	R-squared	0.955	1.000	0.809	0.809			
	No. of observations	4921	4921	4921	4921	4921	4921	4921

1. Estimation period : 1999-2005.
2. FXE denotes Fixed Effect Estimation, LP denotes the estimation following Levinsohn and Petrin (2003), and ENDOG denotes the estimation methodology developed in this paper.
3. In (6) labor is a variable input, whereas in (7) it is a fixed input.
4. *, **, and *** indicate $p < 0.1$, $p < 0.5$, and $p < 0.01$, respectively.

Table 2. Production function estimation

		LP (1)		ENDOG1 (2)		ENDOG2 (3)	
Manufacturing industry	Labor	0.136	***	0.087	***	0.130	***
		(0.007)		(0.020)		(0.022)	
	Capital stock	0.010	***	0.043	***	0.043	***
		(0.000)		(0.000)		(0.008)	
	Material	0.320	**	0.870	***	0.826	***
		(0.128)		(0.042)		(0.024)	
	No. of observations	10713		10713		10713	
Chemical industry	Labor	0.180	***	0.043		0.043	
		(0.024)		(0.137)		(0.067)	
	Capital stock	0.010		0.391	***	0.348	***
		(0.015)		(0.016)		(0.087)	
	Material	0.980	**	0.783	***	0.348	***
		(0.401)		(0.103)		(0.101)	
	No. of observations	1453		1453		1453	
Metal industry	Labor	0.100	***	0.087		0.130	***
		(0.024)		(0.124)		(0.043)	
	Capital stock	0.870	**	0.043	**	0.043	*
		(0.394)		(0.018)		(0.025)	
	Material	0.980	***	0.478	***	0.304	***
		(0.031)		(0.115)		(0.058)	
	No. of observations	625		625		625	
General machinery industry	Labor	0.161	***	0.043		0.174	***
		(0.026)		(0.039)		(0.053)	
	Capital stock	0.190		0.043		0.043	*
		(0.150)		(0.000)		(0.024)	
	Material	0.670	***	0.826		0.826	***
		(0.134)		(0.040)		(0.039)	
	No. of observations	1566					
Electronic and electrical machinery industry	Labor	0.137	***	0.087	***	0.174	***
		(0.035)		(0.027)		(0.037)	
	Capital stock	0.010		0.087	***	0.087	***
		(0.068)		(0.000)		(0.000)	
	Material	0.370	*	0.913	***	0.913	***
		(0.204)		(0.049)		(0.052)	
	No. of observations	1806		1806		1806	
Transportation machinery industry	Labor	0.153	***	0.130	***	0.130	***
		(0.015)		(0.027)		(0.020)	
	Capital stock	0.980	**	0.043	**	0.087	***
		(0.441)		(0.022)		(0.000)	
	Material	0.980	***	0.913	***	0.870	***
		(0.031)		(0.008)		(0.021)	
	No. of observations	931					
Precision machinery industry	Labor	0.142	***	0.130	*	0.130	
		(0.051)		(0.071)		(0.095)	
	Capital stock	0.010		0.087	***	0.174	***
		(0.122)		(0.011)		(0.035)	
	Material	0.180		0.913	***	0.696	***
		(0.304)		(0.105)		(0.086)	
	No. of observations	369		369		369	
Wholesale and retail industry	Labor	0.262		0.174	***	0.261	***
		(0.014)		(0.061)		(0.029)	
	Capital stock	0.010		0.043	**	0.043	
		(0.085)		(0.021)		(0.027)	
	Material	0.290		0.782	***	0.696	***
		(0.263)		(0.049)		(0.037)	
	No. of observations	4510		4510		4510	

1. Estimation period : 1999-2005.
2. The data include only firms which reported R&D expenditure for every year or that reported no R&D expenditure in every year.
3. In (2), labor is a variable input, whereas in (3) it is a fixed input.
4. *, **, and *** indicate $p < 0.1$, $p < 0.5$, and $p < 0.01$, respectively.

Table 3. Summary statistics for productivity indexes

	N	Mean	SD	Min.	Median	Max.
W	9,985	0.768	1.839	-2.517	1.086	5.543
lnTFP1	9,985	0.241	0.199	-0.749	0.209	1.397
lnTFP2	9,985	-0.043	0.173	-1.148	-0.047	0.998

1. Period: 1999-2005.
2. W is the productivity index developed in this paper.
3. lnTFP1 is the productivity index developed by Good et al. (1997) and Aw et al. (1997).
4. lnTFP2 is the cross-sectional productivity index.

Table 4. Correlation between productivity indexes

	Corr. with lnTFP1	Corr. With lnTFP2
Manufacturing industry	0.376	0.084
Chemical industry	0.459	0.471
Metal industry	0.290	0.287
General machinery industry	0.559	0.519
Electronic and electrical machinery industry	0.699	0.680
Transportation machinery industry	0.193	0.159
Precision machinery industry	0.503	0.482
Wholesale and retail industry	0.808	0.800

1. The estimation period is 1999-2005.
2. All values are significant at the 5% significance level.

Table 5. Causes of the productivity bias

	ω			*Labor share*					*Capital share*					*Intermediate input share*				
	Downward biased		Upward biased	Cost share	β_L	Cost share			Cost share	β_K	Cost share			Cost share	β_M	Cost share		
						Downward biased		Upward biased			Downward biased		Upward biased			Downward biased		Upward biased
	(1)		(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)		(10)	(11)	(12)	(13)		(14)
Chemical industry	4.403	**	4.363	0.167	0.043	0.153	***	0.184	0.043	0.348	0.041	***	0.047	0.789	0.391	0.806	***	0.769
	(0.012)		(0.015)	(0.002)		(0.002)		(0.003)	(0.001)		(0.001)		(0.001)	(0.002)		(0.003)		(0.003)
Metal industry	-1.887		-1.896	0.197	0.217	0.206	***	0.187	0.031	0.043	0.032		0.031	0.772	0.913	0.762	***	0.782
	(0.012)		(0.012)	(0.003)		(0.005)		(0.004)	(0.001)		(0.001)		(0.001)	(0.004)		(0.006)		(0.004)
General machinery industry	0.009		0.006	0.201	0.130	0.236	***	0.170	0.028	0.043	0.030	***	0.026	0.770	0.870	0.734	***	0.804
	(0.005)		(0.005)	(0.002)		(0.003)		(0.002)	(0.0005)		(0.001)		(0.001)	(0.002)		(0.003)		(0.003)
Electronic and electrical machinery industry	-1.318		-1.324	0.180	0.130	0.201	***	0.155	0.033	0.087	0.034	**	0.031	0.787	0.913	0.765	***	0.814
	(0.008)		(0.016)	(0.002)		(0.003)		(0.003)	(0.001)		(0.001)		(0.001)	(0.002)		(0.003)		(0.003)
Transportation machinery industry	-0.802		-0.792	0.181	0.130	0.185		0.177	0.044	0.087	0.046	***	0.041	0.776	0.870	0.769	**	0.782
	(0.007)		(0.008)	(0.002)		(0.003)		(0.003)	(0.001)		(0.001)		(0.001)	(0.003)		(0.004)		(0.004)
Precision machinery industry	2.542		2.539	0.204	0.130	0.176	***	0.241	0.030	0.348	0.028	***	0.034	0.766	0.435	0.796	***	0.726
	(0.020)		(0.024)	(0.004)		(0.005)		(0.006)	(0.001)		(0.001)		(0.002)	(0.005)		(0.006)		(0.007)
Wholesale and retail industry	1.284	***	1.318	0.392	0.261	0.422	***	0.367	0.031	0.043	0.031		0.031	0.576	0.696	0.547	***	0.601
	(0.006)		(0.004)	(0.002)		(0.004)		(0.002)	(0.0004)		(0.0007)		(0.0005)	(0.002)		(0.004)		(0.002)

1. The estimation period is 1999-2005.
2. Numbers shown are the group means, with the standard deviation for the group shown in parentheses.
3. ***, **, and * indicate that the means of the two groups significantly differ at the 1%, 5%, and 10% significance level, respectively.

Table 6. R&D returns

	Obs.	**Mean**	**Median**	Std. dev.	Min.	5%	25%	75%	95%	Max.
Manufacturing industry	6,461	**0.05**	**0.04**	0.36	-1.78	-0.76	-0.05	0.22	0.58	1.19
Chemical industry	1,438	**0.35**	**0.38**	0.21	-0.49	-0.08	0.25	0.49	0.63	0.74
Metal industry	586	**-0.81**	**-0.78**	0.19	-1.78	-1.16	-0.92	-0.69	-0.52	-0.30
General machinery industry	1,469	**0.18**	**0.18**	0.03	0.05	0.14	0.16	0.20	0.24	0.28
Electronic and electrical machinery industry	1,706	**-0.01**	**-0.01**	0.03	-0.14	-0.04	-0.02	0.00	0.05	0.11
Transportation machinery industry	912	**-0.16**	**-0.17**	0.06	-0.34	-0.27	-0.20	-0.13	-0.06	0.03
Precision machinery industry	350	**0.56**	**0.58**	0.22	0.03	0.06	0.49	0.69	0.83	1.19
Wholesale and retail industry	754	**-0.02**	**-0.02**	0.04	-0.16	-0.09	-0.04	0.00	0.05	0.14

1. The estimation period is 1999-2005.

Table 7. R&D return and firm characteristics

	Manufacturing industry	Chemical industry	Metal industry	General machinery industry	Electronic and electrical machinery industry	Transportation machinery industry	Precision machinery industry	Wholesale and retail industry
Log(Sales)	-0.021 ***	0.099 ***	-0.094 ***	0.017 ***	0.013 ***	-0.033 ***	0.050 ***	-0.013 ***
	(0.004)	(0.003)	(0.007)	(0.000)	(0.000)	(0.001)	(0.009)	(0.001)
Return on assets	0.552 ***	0.290 ***	0.043	-0.005	-0.035 ***	0.060 *	0.440 ***	0.056 **
	(0.092)	(0.047)	(0.141)	(0.008)	(0.007)	(0.032)	(0.115)	(0.022)
Debt/Total assets	-0.084 ***	-0.107 ***	0.092 ***	-0.018 ***	-0.014 ***	0.032 ***	0.054	0.000
	(0.023)	(0.011)	(0.027)	(0.002)	(0.002)	(0.006)	(0.039)	(0.006)
GOV	-22.500 ***	-11.700 ***	-0.219	0.608	-0.495	5.960 ***	-48.300	0.472
	(3.278)	(4.234)	(1.395)	(0.977)	(0.795)	(1.174)	(83.089)	(0.386)
FIN	0.475 ***	0.018	-0.281 ***	0.017 ***	0.006	-0.004	-0.164 ***	-0.040 ***
	(0.037)	(0.017)	(0.057)	(0.003)	(0.004)	(0.009)	(0.063)	(0.009)
STC	0.540 *	-0.015	2.430 ***	-0.154 ***	0.021	0.238 **	-0.684 **	-0.033
	(0.323)	(0.182)	(0.410)	(0.026)	(0.030)	(0.098)	(0.332)	(0.041)
FRN	0.231 ***	0.007	-0.167 **	0.011 **	0.017 ***	-0.034 ***	0.131 *	-0.060 ***
	(0.046)	(0.022)	(0.074)	(0.005)	(0.004)	(0.009)	(0.075)	(0.012)
PER	-0.026	0.065 ***	0.062 *	0.007 ***	0.005 *	-0.006	0.108 **	0.023 ***
	(0.029)	(0.014)	(0.034)	(0.003)	(0.003)	(0.007)	(0.050)	(0.007)
Constant	0.308 ***	-1.370 ***	0.731 ***	-0.099 ***	-0.236 ***	0.406 ***	-0.345 **	0.192 ***
	(0.074)	(0.042)	(0.118)	(0.007)	(0.006)	(0.017)	(0.159)	(0.019)
R-squared	0.073	0.877	0.558	0.763	0.646	0.770	0.705	0.420
No. of observations	6493	1441	587	1488	1711	914	352	757

1. The estimation period is 1999-2005.
2. Numbers in parentheses are standard errors.
3. *, **, and *** indicate $p < 0.1$, $p < 0.5$, and $p < 0.01$, respectively.
4. Year dummies are included in the estimation but not reported here.
5. GOV, FIN, STC, FRN, and PER denote the ratios of shares owned by the government, financial firms, incorporated (non-financial) firms, foreign firms, and private investor

Table 8. Tobin's Q and R&D returns

	Manufacturing industry		Chemical industry		Metal industry		General machinery industry		Electronic and electrical machinery industry		Transportation machinery industry		Precision machinery industry		Wholesale and retail industry	
R&D returns	0.165	***	0.689	***	0.055		2.570	***	1.670	***	-1.260	***	0.367		0.047	
	(0.014)		(0.069)		(0.062)		(0.355)		(0.354)		(0.150)		(0.236)		(0.349)	
Return on assets	2.440	***	3.480	***	1.500	***	1.510	***	2.560	***	2.720	***	2.600	***	2.760	***
	(0.115)		(0.252)		(0.301)		(0.231)		(0.233)		(0.306)		(0.533)		(0.245)	
Constant	0.654	***	0.369	***	0.608	***	0.228	***	0.752	***	0.304	***	0.569	***	0.658	***
	(0.006)		(0.025)		(0.054)		(0.065)		(0.012)		(0.027)		(0.126)		(0.018)	
R-squared	0.164		0.320		0.133		0.153		0.159		0.284		0.179		0.227	
No. of observations	4377		1002		368		1020		1172		584		231		519	

1. The estimation period is 1999-2005.
2. Numbers in parentheses are standard errors.
3. *, **, and *** indicate p < 0.1, p < 0.5, and p < 0.01, respectively.
4. Year dummies are included in the estimation but not reported here.