R and D Performance in Chinese Industry[1]

Gary H. Jefferson
Graduate School of International Economics and Finance
Brandeis University, Waltham, MA
Jefferson@brandeis.edu

Bai Huamao
Graduate School of International Economics and Finance
Brandeis University, Waltham, MA

Guan Xiaojing
Department of Population, Social, and Technology Statistics
National Bureau of Statistics

Yu Xiaoyun
Department of Industrial and Transportation Statistics
National Bureau of Statistics
Beijing, China

December 28, 2002 (final)

Forthcoming in *Economics of Innovation and New Technology*

ABSTRACT

This research, which investigates a set of fundamental relationships in the R&D literature, is based on an unusually rich set of panel data covering the population of China's large and medium-size manufacturing enterprises. Using a recursive three-equation system, we investigate the determinants of firm-level R&D intensity, the process of knowledge production, and the impact of innovation on firm performance. Several results stand out. Overall, the statistical relationships within the model are surprisingly robust, including the contributions of R&D expenditure to new product innovation, productivity, and profitability. The roles of firm size, market concentration, and profitability in driving R&D effort parallel those found in the U.S. literature. We find that new product innovation accounts for approximately 12 percent of the total returns to R&D. Also, returns to industrial R&D in China appear to be at least three to four times the returns to fixed production assets.

---

[1] We appreciate the assistance of Messers Ma Jingkui and Xu Jianyi of China's National Bureau of Statistics. Their collaborative made this research possible. We also appreciate the support of the Henry Luce Foundation, the National Science Foundation (grant #SES-9905259), the U.S. Department of Energy's Biological and Environmental Research Program (contract # DE-FG02-00ER63030), the William Davidson Institute at the University of Michigan, and the Rockefeller Center at Dartmouth College. The helpful suggestions of Blake LeBaron, Albert Hu, Adam Jaffe, Jacques Mairesse, and Tom Rawski are also very much appreciated as are comments received during seminars at Beijing University, MIT, the National Bureau of Economic Research, and Tsinghua University.

1. Introduction

During the last two decades of the 20th century China's economic reforms spurred impressive rates of economic growth. Much of this growth, however, has arisen from transitional gains in allocative efficiency that has resulted from the initial stages of market and ownership reform. In order to continue along the path of convergence with the world's advanced economies, China's manufacturing sector must learn how to deploy and utilize R&D resources effectively. Over the long-term, China's economic performance will depend on its ability to innovate, acquire and diffuse new technologies, and eventually to become a leading contributor to the world's stock of industrial knowledge.

This paper uses an extraordinary set of firm-level data to investigate three questions relating to research and development in China's manufacturing sector. These are also the three classic questions that frame the considerable research on firm-level R&D in the advanced industrial countries: (i) what determines variations in R&D effort across enterprises?; (ii) how does R&D effort translate into R&D output, i.e. can we identify knowledge production functions at the firm level?; and (iii) how do levels of R&D activity, measured in terms of R&D inputs and outputs, affect firm-level productivity and profitability?

To investigate these questions, we use a panel of firm-level data that spans China's approximately 20,000 large and medium-size manufacturing enterprises that are grouped according to seven different ownership types. While measures of R&D effort in China's small-size enterprise sector are incomplete, we anticipate that, like the U.S.,

China's larger enterprises, which account for approximately one-third of total manufacturing sales, sponsor the vast majority of that country's manufacturing R&D activity.[2]

This paper is organized as follows. In the following section we outline the three-equation model that organizes our research strategy. Section 3 provides a statistical profile of the large and medium-size enterprises that comprise our data set and the nature of their R&D activities. Sections 4, 5, and 6 outline the estimation equations for R&D intensity, knowledge production, and firm performance respectively. Section 7 describes the way in which we constructed the set of panel data that we use for our estimation work; using these data, in Section 8 we present the estimation results for the three-equation model. Deriving a reduced form from the knowledge and performance equations that link R&D expenditure directly with firm performance, Section 9 presents results for estimates of the total returns to R&D expenditure in Chinese industry and compares these with returns to investment in fixed capital. Finally, Section 10 identifies the fraction of total returns to R&D associated with the new product innovation channel. Section 11 summarizes our conclusions. Throughout, we attempt to relate our results to established findings in the innovation literature based on U.S. manufacturing data.

2. Modeling R&D

Economists in the U.S. have developed an extensive literature that investigates the impact of R&D on firm performance. While this literature as a whole focuses on the

[2] According to Hay and Morris (1991), in the U.S.A. in 1970 the largest 100 employers accounted for 79 percent of total manufacturing R&D expenditure.

three questions posed above, efforts to examine all three of these issues simultaneously within an integrated model of the innovation process are relatively recent.[3]

We construct and estimate a model, which consists of the following three equations:

$$\text{R\&D intensity: } (\text{RDexp/SALES})(t-1) = f_1[X(t-2)]\varepsilon_1, \quad (1a)$$

$$\text{R\&D output: } \text{RDout/SALES}(t) = f_2[(\text{RDexp/SALES})(t-1), Y(t)]\varepsilon_2, \quad (1b)$$

$$\text{Performance: } \text{PERF}(t) = f_3[(\text{RDout/SALES})(t), Z(t)]\varepsilon_3. \quad (1c)$$

In this recursive 3-equation model, spanning three periods, R&D intensity is a function of conditions and expectations that are determined by prior period variables, including firm sales, profits, and industry concentration, summarized by the vector X. In estimating the contribution of these factors to R&D intensity, we control for the separate effects of ownership status and industry affiliation. We use as our measure of innovation activity technological innovation expenditure (*jishu kaifa zhichu*), a somewhat broader measure of innovation activity than that encompassed by the more conventional counterpart measure of R&D expenditure.[4] Our measure of R&D intensity is the ratio of technological innovation expenditure to sales, i.e. RDexp/SALES.

R&D output, measured here as the ratio of new product sales to total sales, is driven by last-period R&D intensity, i.e. [RDexp/SALES](t-1), as well as age. We also investigate the impact of technological opportunity, proxied by industry classifications,

[3] See Crepon, Duguet, and Mairesse (1998), van Leeuwen and Klomp (2001), and Loof and Heshmati (2001).
[4] See the *Frascoti manual* at www.oecd.org/EN/about/0,,EN-about-571-nondirectorate-no-no-no-18,00.html , p. 5 for definitions of the respective scopes of R&D and technological innovation.

and incentive structure, proxied by ownership structure, on the efficiency of knowledge production.

Finally, firm performance, measured by productivity and profitability, is driven by available R&D outputs, which serve as inputs to the production process, and by the usual factor inputs that are summarized by the vector Z. We again include ownership classification and industry type to distinguish their independent effects on firm performance. Using measures of output and profit, we estimate a productivity function and a profit function.

Within our three-equation model, research and development provides the bridge between the set of technical possibilities that the firm's manager observes in the wider world and the technologies that are realized within the firm. While innovation outside the firm will gradually enlarge the set of product and process possibilities that are embodied in purchased equipment and hired-in labor, work by Mowery (1983) and Basant and Fikkert (1996) demonstrate that the firm's own investment in R&D will accelerate these spillover effects. With the promise of higher future returns from new products and process technologies, the manager solves a first-order condition, which equates the expected stream of marginal benefits from R&D effort to the marginal cost of R&D.[5] Specifically, the agent chooses an optimal stock of R&D and the net adjustment to that stock in each period. These choices are based on his familiarity with the technologies implicit in the stochastic knowledge production function summarized in equation (1b) and the performance equation summarized in equation (1c), as well as the

[5] We assume an optimizing problem similar to the formulation of Cohen and Klepper (1996), who explore the role of firm size and R&D intensity within a profit-maximizing framework. In addition to size, we hypothesize that market conditions and other firm-specific characteristics also shape the R&D resource allocation decision.

reduced form equation, described below, which captures the total direct effect of R&D effort on performance.

Estimates from this three-equation model will enable us to calculate and compare the returns to R&D resources across firm size and ownership classifications. Using the results from equations (1a) – (1c), we can estimate the impact of R&D activity on enterprise performance (PERF) as it operates through new product innovation, i.e. $(dPERF/dRDexp)_{NP} = (\partial PERF/\partial NP)(\partial NP/\partial RDexp)$. In addition, by substituting equation (1b) into (1c), we estimate a reduced form equation, which examines directly the total impact of R&D inputs on performance, i.e.$[dPERF/dRDexp]_T$. Because R&D can be expected to operate through multiple channels that are unlikely to be captured fully by new product sales, such as quality improvements and process innovation, we anticipate that the total impact (T) of R&D will exceed the combined impact of the new product innovation (NP) channel alone, i.e. $[dPERF/dRDexp]_T > [dPERF/dRDexp]_{NP}$. We test this hypothesis in Section 10.

3. China's R&D Performers

Our data set is drawn from a sample of nearly 20,000 large and medium-size manufacturing enterprises (LMEs).[6] As shown in Table 1, in 1999, the most recent year in our set of panel data, approximately one-third of the population of LMEs were classified as "large-size". These firms, on average, report a volume of sales that is six times as large as that of the medium-size enterprises, net value of fixed assets (NVFA)

[6] The data are collected and tabulated annually by China's National Bureau of Statistics (NBS).

that is eight times as large, and a work force that is four times as large. Profitability – here measured as the ratio of the difference between sales revenue and the production cost of units sold to total sales – is somewhat higher for the large-size firms.

Table 2 reports the means for available measures of technical innovation expenditures and employment as well as a key innovation output, i.e. new product sales and patent applications. These statistics include means for the full population of LMEs, as well as for only those enterprises that report positive counts of R&D inputs or outputs. As we will see, a problem with the data set for the full population is that for many enterprises that did not report their data, the statistical authorities have converted blank observations to zeros. The data for the full population, therefore, cannot distinguish between non-performers and non-reporters. According to Table 2, in 1999, 14,037 enterprises, more than 70 percent of the total population, reported a positive level of R&D activity for at least one of the four R&D measures reported in that table. To avoid the ambiguity concerning the meaning of a zero, in Table 2, we also report the means for only those enterprises reporting positive counts for each the R&D measures.

As shown in Table 2, the intensity of R&D activity, for both inputs and outputs, rose substantially from 1995 to 1999. Whether we compare the full population over time or only those firms reporting positive counts for each category of R&D activity, most of the magnitudes increase by a factor of nearly two or more. This rise in overall intensity transpires notwithstanding the decline in the total number of enterprises reporting positive levels of activity for R&D personnel and new product sales.

Using the full data set for 1999 and probit analysis, Table 3 distinguishes between the characteristics of R&D performers (designated “1”) and other firms (designated “0”)

consisting of some unknown combination of R&D non-performers and non-reporters. The reference intercept in the regression is the sub-sample of medium-size, state-owned enterprise, in industry 31, i.e. non-metal mineral products. The likelihood of a firm in this reference group being an R&D performer is 51.9 percent. The results show that sales, capital intensity, and designation as a "large-size" enterprise all significantly increase the probability of being an R&D performer. Furthermore, controlling for these factors and industry affiliation, R&D performers are concentrated among state owned and shareholder owned enterprises. The foreign and overseas (HKT) sectors include a low incidence of R&D performers, which may reflect the fact that many firms in these sectors draw on innovations from their parent companies. Finally, high concentrations of R&D activity are found in the instrument (42), special purpose equipment (36), ordinary machinery (35), and medical and pharmaceutical products sectors. The lowest concentrations of R&D activity are reported in the printing (23), garment (18), and furniture manufacturing (20) sectors.

In principle, we can correct for the selection bias implicit in the narrowing of our sample from nearly 20,000 to a fraction of that population. However, because the data do not allow us to distinguish among R&D performers, who do not report, non-performers who do report, and non-performers, who do not report, we have no basis for interpreting a selection bias correction.[7] The only selection bias issue that we investigate is that associated with the omission of loss-making enterprises from our sample.

[7] For example, if all of the omitted enterprises, for which we have data for 1997-99, are non-reporting R&D performers and, moreover, they are identical to the reporting R&D performers (that constitute our sample), then no correction for selection bias is required. On the other hand, if these firms are R&D performers that are reporting (zero values), then we can implement a correction for selection bias. Unfortunately, we have no basis for identifying the ratio of non-reporting performers and reporting non-performers.

4. R&D Effort

R&D effort is the outcome of the interaction of factors that condition the demand for and supply of R&D resources. Within the context of the model outlined in Section 2, a manager is assumed in period t-2 (i.e. 1997) to be setting next period's R&D intensity, conditional on the firm's expected sales, driven by current sales in t-2, and the level of market concentration in the relevant four-digit industry, i.e. (CR2)(t-2). Industry market concentration, measured as the share of the combined sales of the two largest enterprises in total industry sales, varies, for instance, due to differences in barriers to entry, product differentiation, and past innovation. The manager's choice of R&D effort is shaped by profitability (PROF) in t-2, since retained earnings are an important source of innovation finance, as well as by the set of technological opportunities that are intrinsic in the manager's industry (IND). The incentive structure and the regulatory environment, which, in China, may be associated with ownership structure (OWN), also affect the manager's demand for R&D resources. The supply curve is determined partially by the firm's profitability and retained earnings that can be invested in the R&D operation. Where R&D depends on outside financing, the firm's financial condition, ownership structure, and industry affiliation will also bear on the supply of bank finance or government funding.

We summarize a reduced form for R&D effort as:

$$\ln(RDexp/SALES_{-1} = \alpha_0 + \alpha_1 \ln SALES_{-2} + \alpha_2 \ln PROF_{-2} + \alpha_3 \ln CR2_{-2}$$

$$+ \alpha_4(\ln CR2_{-2})^2 + \alpha_5 \ln(RDexp/SALES)_{-2} + \Sigma\alpha_j IND + \Sigma\alpha_i OWN + \varepsilon_1. \quad (2)$$

R&D effort may also be conditioned by the level of R&D effort in the last period. The firm's ability to utilize R&D resources effectively, even in the presence of favorable movements in firm size, profits and other drivers of R&D effort, may be limited by the firm's established R&D capabilities. The robustness of the coefficient on lagged R&D effort measures the persistence of R&D intensity; as such it is a useful measure of R&D flow as a proxy for R&D stock, the preferred measure of R&D effort. We estimate equation (2) with and without a lag in R&D intensity. The concentration ratio (CR2), constructed from the data set for each of China's 552 four-digit industries, tests the competing perspectives of Schumpeter (1942), who emphasizes the functions of monopoly power in driving technological progress, and Arrow (1962), who demonstrates the advantages of competitive markets to innovation.

In the U.S. economy, retained earnings is a major funding source of R&D activity (Grabowski, 1968). In China, where enterprises face particularly severe liquidity constraints, the importance of retained earnings may be quite pronounced. On the other hand, as competition and the legal system evolve in China, poor profit performance may motivate R&D effort, including R&D subsidies from the state, to avoid chronic losses and bankruptcy.

The demand for R&D resources will vary across industries depending on the degree of technological opportunities intrinsic to the industry. Comanor and Scherer

(1969), for example, argue that a pharmaceutical innovation is relatively easy to patent; it is much harder for an invention like a ballpoint pen to be patented in a way that will exclude competition. New product innovation may also be limited in industries that produce relatively homogeneous goods, such as petroleum and natural gas production, mineral mining, and tobacco processing.

Finally, ownership structure may matter in multiple ways. The entrepreneur's incentive structure and time horizon will be affected by the system of ownership. Also, ownership may influence the availability of R&D resources, such as government subsidies that are more readily available to the state sector. Finally, the opportunity for R&D decision-makers to capture the returns to innovation may vary across ownership types due to differences in tax systems and enforcement of intellectual property rights.

5. Knowledge Production Functions

In recent decades, the knowledge production function has emerged as an economic concept that links R&D outputs to R&D inputs.[8] The survey form, used by China's National Bureau of Statistics to collect the data, defines new products as those that were introduced in production during the past three years. We represent the relationship between R&D inputs and R&D using the following intensive functional form:

[8] For a review of research in this area, see Griliches (1998). The only application of the knowledge production function to Chinese industry with which we are familiar is Hu (1998).

$$\ln(NP/SALES) = \alpha_0 + \alpha_1 \ln(RDexp/SALES)_{-1} + \alpha_2[\ln(RDexp/SALES)*\ln SALES]_{-1}$$
$$+ \alpha_3 \ln AGE + \Sigma\alpha_j IND + \Sigma\alpha_i OWN + \varepsilon_2 \qquad (3)$$

The R&D variable enters in intensive form for the following reason. As an input, R&D expenditure includes expenditures on labor, capital, and intermediate inputs. While double counting would not be a problem in equation (3), since it does not include other factor inputs on the right hand side, it will prove to be a problem in estimating the reduced form production and profit functions in Section 10, in which performance is driven by both R&D inputs and by capital, labor, and intermediate inputs whose measures include inputs to measured R&D expenditure. To avoid double counting in the reduced form, which is derived by substituting equation (3), the R&D output equation, into equation (7), the productivity equation, we construct the variable RDexp/SALES to represent the R&D intensity of the {K,L,M} vector of inputs.[9]

A central issue in the estimation of knowledge production functions is the role that firm size plays in determining the efficiency of knowledge production. We explicitly test this scale hypothesis through the inclusion of a cross-product, $\ln(RDexp/SALES)*\ln SALES]$, in the estimation equation. Based on this functional form, the marginal product of R&D inputs with respect to R&D output is measured as:

$$\partial NP/\partial RDexp_{-1} \cong [\alpha_1 + \alpha_2(\ln SALES)_{-1}](NP/RDexp_{-1}) \qquad (4)$$

[9] Schankerman (1981) discusses some of the estimation implications of double counting R&D inputs.

Values of $\alpha_2 \neq 0$ imply non-neutral scale effects. Positive scale effects may arise from internal economies of scope, say as between learning-by-doing associated with production scale and R&D, or they may capture external marketing capabilities that make it easier to launch new products. Alternatively, negative scale effects may reflect problems of coordination or disincentive effects associated with scale.

We include in our knowledge production function a test for the role of experience, i.e. the age of the firm (AGE), which may serve to enhance through learning effects or impair, say through organizational sclerosis, a firm's R&D capabilities. Finally, different ownership forms may imply different incentive structures that impact on R&D efficiency. Also, within different industries, the degree of difficulty is likely to vary in terms of the requisite volume of R&D inputs that are needed to produce comparable measures of new products.

We employ a commonly used measure of R&D output: the ratios of new product sales to total sales. To allow for a broader measure of the impact of R&D on firm performance, such as process innovations and improvements in product quality that do not qualify as "new products," we derive and estimate in Section 10 a reduced form version of equations 1(b) and 1(c). Even if new product sales turn out be an incidental aspect of the innovation process, i.e. they are statistically insignificant in equations 1(b) or 1(c); the reduced form estimates may still reveal that R&D effort exerts a significant impact on firm performance through channels other than the development of new products.

6. Performance

We investigate the impact of R&D on two performance measures: productivity and profitability. In principle, in a perfectly competitive setting, with uniform technologies, a firm's relative productivity and profitability should perfectly correspond. In practice, they may not for various reasons. Although we use industry dummies to capture technology differences, within two-digit industry classifications, technologies are likely to vary significantly. Both within and across industries, the markup of price over marginal cost will vary with market concentration and with ease of entry in both input and product markets. Differences in innovation intensity, predicted by equation (2), are likely to contribute to differences in market power and price markups. Furthermore, differences in productivity, which we estimate from available input and output data, will not correlate fully with differences in profitability reported by the individual enterprises. For these reasons – heterogeneous technologies, imperfect markets, and measurement error – we investigate separately the impact of R&D on productivity and profitability.

Productivity. Production is modeled as a Cobb-Douglas technology in which gross output (Q) is driven by inputs of fixed capital (K), labor (L), and intermediate inputs (M):

$$\ln Q = a + \alpha_K \ln K + \alpha_L \ln L + \alpha_M \ln M + \varepsilon_3 \qquad (5)$$

Productivity differences arise from differences in R&D performance, ownership, and industry-specific technologies:

$$a = a_0 + \theta \ln(NP/SALES) + \Sigma\alpha_i OWN + \Sigma\alpha_j IND + \varepsilon_4. \qquad (6)$$

Substituting equation (6) into equation (5), the third estimation equation in the model becomes:

$$\ln Q = a_0 + \theta \ln(NP/SALES) + \alpha_K \ln K + \alpha_L \ln L + \alpha_M \ln M + \Sigma\alpha_i OWN + \Sigma\alpha_j IND + \varepsilon_5, \qquad (7)$$

where $\varepsilon_5 = \varepsilon_3 + \varepsilon_4$.

Profitability: To investigate the impact of R&D innovation on profitability, we start with a restricted profit function, which includes only the fixed or quasi-fixed factors (i.e. capital and labor). Substituting equation (6) into the restricted profit function yields:

$$\ln PROF = b_0 + \theta \ln(NP/SALES) + \gamma_K \ln K + \gamma_L \ln L + \Sigma\alpha_j IND + \Sigma\alpha_i OWN + \varepsilon_6. \qquad (8)$$

We justify the omission of a term for intermediate inputs (i.e. M), i.e. the restriction $\alpha_M = 0$, based on the assumption that intermediate input markets for China's LMEs are substantially competitive and, unlike capital and labor, these inputs are flexible. We support this assumption with results from Jefferson et al. (2000), who find that, unlike expenditures on capital and labor, the marginal products to intermediate inputs within and across ownership categories in China's industrial sector are very close

to unity. That is, one additional dollar of expenditure on intermediate inputs raises output by one dollar and therefore apparently has no effect on profit.

7. The Data and Econometric Issues

Our data set covers the years 1995-1999. In principle, we should use all the data, or a balanced set of data, to estimate our model. Using five years of panel data, we can (i) construct a stock of R&D capital and/or (ii) incorporate time series analysis into our estimation. In this paper, we use only the recent three years of the data: i.e. 1997, 1998 and 1999.[10] We do this for two reasons. These are:

- The number of enterprises for which we would be able to construct a meaningful stock of R&D capital over a five-year period is severely limited by the large number of enterprises for which observations are missing in one or more years.
- The further back we go in time, the more homogeneous the ownership structure of the data set. Many of the non-state owned enterprises that are present in the data set during 1997-1999 were not present during 1995-1996.

Table 4 shows the cross sectional distribution of our panel by size and ownership using data balanced over 1997-1999. The table reports the number of R&D performers in each size and ownership classification, i.e. those that report positive counts for each of the four R&D variables. The panel also identifies the number of observations for which profit is non-negative.

In addition to being included in the panel for each of the three years – 1997 to 1999 – for an enterprise to be included in our estimation sample, firms must conform to two additional criteria. These are: (i) be designated as an "R&D performer" by reporting a positive level of R&D activity for at least one of the four R&D measures in each of the three years and (ii) report non-negative profits in 1999. By using the criterion of at least one R&D measure greater than zero for inclusion, we assume that if a designated R&D performer reports zero for any one of the other three R&D measures, then that zero can be interpreted as a true measure.

The bottom row in Table 4 identifies the size and ownership distribution of the subsample that we use for estimating our R&D model. If we were to use only enterprises in 1999 that satisfy our condition for being R&D performers, then our sample for that year would shrink to 14,037 enterprises, as shown in the last column of Table 2. Balancing our panel set of R&D performers for three years – 1997, 1998 and 1999 – reduces our sample size to 5,740 enterprises. As shown in Table 4, when we further impose the restriction that firms cannot be loss-making enterprises in 1999, the sample shrinks to 5,451 enterprises.

A second issue that arises in relating our data to the estimation equations is the role of prices. Since we are using a single cross section, we are generally able to avoid the problem of constant versus current prices. In estimating the production function, we use a current price measure of gross output. Intermediate inputs, R&D expenditure, new product sales, and total sales are also denominated in nominal prices. Labor is measured in physical units. The capital stock embodies a series of vintages, each associated with a contemporaneous price. Ideally, we would have a current market valuation of each

[10] A statistical description of the data can be found in Jefferson et al (forthcoming).

firm's fixed assets, but these data are not reported. Our use of industry and ownership dummies serves to capture some of the differences in the price structure of industrial fixed assets.

According to our selection criteria, all 5451 enterprises in our sample qualify as R&D performers. However, in one or more years some enterprises may report a zero level of activity for certain variables, namely R&D expenditure, new product sales, or profit. Because the estimation equations in our model are typically in log-log form, logarithmic values of these zero observations are indeterminate. In order to maintain a consistent enterprise sample across the equations in our model, we include the zero observations by transforming them to a value $X_m + (0.0001)k_m$, where X_m represents a zero observation for the m^{th} variable and k_m represents a scalar adjustment to the arbitrarily chosen mark up of 0.0001. By including a dummy for each of the $X_m = 0$ observations, the regression procedure implicitly chooses a value of k_m that recenters the $\ln(0.0001k_m)$ observations around the regression line that has been defined by the non-zero observations.[11]

Using a set of panel data that spans three years, our three-equation model embodies a time structure that, in principle, avoids the problem of simultaneity along the time dimension. In fact, however, due to the inter-temporal correlation of certain key variables, such as profit, output, sales, and R&D expenditures, associated with fixed firm effects, the data may be driven by simultaneous relationships that persist over time within

[11] While dummies for the zero observations associated with explanatory variables are recentered along the vertical axis so as to leave the regression unaffected, dummies for zero observations associated with the dependent variable form a horizontal line that is shifted up to center horizontally on the regression line that has been fit to the non-zero observations. The effect of the horizontal row of zero observations on the dependent variable will be to reduce the slope estimate of the regression. This is to be expected, since these zero observations represent the case in which a positive level of input, say R&D expenditure, elicits no output, say in the form of additional new product sales or profit.

our recursive model. Where appropriate and feasible, we use instrumental variables (IV) estimators to minimize estimation bias. In the following section we discuss the endogeneity issues associated with each of the regressions.

Finally, our data include 23 ownership classifications. For purposes of estimation, we combine these into seven classifications, which correspond to the more commonly used set of Chinese industry ownership classifications. These are: state-owned enterprises (SOE), collective-owned enterprises (COE), foreign-invested enterprises (FIE), Hong Kong, Macao, and Taiwan (HKT) or overseas-owned enterprises, shareholding enterprises (STK), privately-owned enterprises (PRI), and other (OTH) forms of domestic ownership. While our data set classifies each enterprise according to one of 552 four-digit enterprises, we aggregate these into the 29 two-digit industries shown in Annex I.

In the regression analysis that follows, the reference intercepts are represented by state-owned enterprises and by nonmetal mineral products (31). Coefficients on dummy variables that are statistically significant indicate significant different – larger or smaller – of the dummied variables relative to these reference categories.

8. Estimation Results

We estimate the R&D intensity equation (2), the knowledge production function (3), the productivity equation (7), and the profitability equation (8). The key results are outlined below.

R&D intensity. In principle, since we have designed a recursive model in which variables in time t-2 drive R&D intensity in t-1, equation (2) should be free of

endogeneity. Our concern, however, is that in practice our panel may exhibit persistent firm effects, so as to cause shocks to sales to be correlated over time. Because of our concern that firm-specific shocks to sales may create spurious correlation between the dependent variable ln(RDexp/SALES)(98) and the SALES(97) variable on the right-hand side, we estimate the regression using instrumental variables for SALES(97). These IVs, consisting of the size dummies,[12] as well as the ownership and industry dummies, lagged one year, are unambiguously exogenous.

Table 5 presents three sets of results. Each includes an instrumental variable for SALES(97), since SALES appears on both the left and right hand side of the equation. We estimate the R&D intensity function with and without the industry dummies. A comparison of the results, shown in the first and second columns of Table 5, indicate that two of the estimated coefficients change significantly as a result of the inclusion of the two digit industry variables. The set of results without the industry dummies shows an elasticity of R&D intensity with respect to sales of -0.201, which indicates a statistically significant tendency for large-size firms to exhibit lower R&D intensities than medium-size firms. Also, in the version without industry dummies, the industry concentration variables enter robustly, showing that R&D intensity rises with industry concentration. Cohen and Levin (1989) conclude that "the majority of studies (in the U.S.) that examine the relationship between market concentration and R&D have found a positive relationship" (p. 1075). Our result is consistent with these findings.

The estimates in the second column of Table 5 that include the industry dummies differ from those in the first column in two significant respects. The first is that the

[12] The size variable includes five categories – three for large-size firms and two for medium-size firms.

coefficient on sales turns positive but insignificant. In their review of the relationship between firm size and R&D intensity in the U.S., Cohen and Levin (1989) conclude: "the most notable feature of the considerable body of empirical research on the relationship between firm size and innovation is its inconclusiveness" (p. 1069). Our result in which we find in Chinese industry an historic bias in R&D intensity toward medium-size firms that disappears when we add in the industry effects firms easily falls within the boundaries of U.S. experience. The second change that results from inclusion of the industry dummies is that estimates of the industry concentration ratios, previously robust, turn insignificant. This result in which robust estimates of industry concentration lose their significance when industry effects are included is consistent with research results reported by Scherer (1987), Scott (1984), and others.[13]

The results shown in column (3) of Table 5 include both industry dummies and the lagged value of RDE/sales. Estimates of the sales and industry concentration variables continue to be insignificant. The notable change relative to columns (1) and (2) is that estimates of the profit variable turn from negative to positive. It appears that whereas, controlling for sales, profit exhibits a negative historical association with R&D intensity, when we focus on the increment to R&D intensity from 1997 to 1998 shown in column (3), profit becomes positively associated with R&D intensity. This finding of a positive association between profit and increments to R&D intensity is consistent with Grabowski (1968), who found in U.S. manufacturing data that profits lagged by one year were a significant determinant of R&D expenditure at the firm level. Given the declining role of the state in financing R&D as well as the generally risky nature of the R&D operation, the difficulty of monitoring R&D activity externally, and the inefficiency of

[13] See Cohen and Levin (1989), p. 1077.

capital markets in China, our finding regarding the increasing importance of profits financing R&D is not unexpected.

In addition, we find that:

- Without industry controls, R&D intensity within the collective sector lags behind other ownership forms. Foreign and overseas firms exhibit relatively high R&D intensities. However, these ownership effects largely disappear with the inclusion of the industry dummies.
- Levin et al (1985) identified key characteristics of industries that exhibit high levels of R&D intensity. These include: (i) dependence on basic science, (ii) dependence on technology supplied by upstream suppliers, and (iii) effectiveness of mechanisms to prevent imitation. Surprisingly, once we control for sales, profit, industry concentration, and ownership, the industry dummies do not show much variation in R&D intensity across industries.
- R&D effort, lagged one period (column 3), enters robustly, which implies a high degree of persistence of R&D effort. We expect, therefore, that recent R&D resource flows provide a reasonable proxy for R&D stock, the more appropriate measure of R&D input. As we show later in the paper, estimates of the elasticity of profit with respect to R&D input are unaffected by the creation of a two-year stock of R&D.

Knowledge production. The estimation results for the creation of new product sales are shown in Table 6. By including the cross-product of ln(RDexp/SALES) and ln(SALES), we test for scale neutrality. The negative estimate on the cross product shows the elasticity of new product sales with respect to R&D expenditure declining with

firm size. In Table 11, however, we see that the marginal product of new product sales is only marginally smaller for large-size enterprise. This result implies that for larger firms the average ratio of NPsales/RDexp used to compute the marginal productivity is larger than for medium-size firms, thereby substantially offsetting the somewhat smaller estimated elasticity for larger firms.

In addition, our findings include the following:

- AGE, as a proxy for experience, is consistently insignificant. It may be that age simply does not matter; alternatively, since the older enterprises are typically SOEs and the foreign and overseas enterprises are newer, these systematic age differences may be captured by ownership effects.
- In the creation of new product sales, the equipment and machinery industries (i.e. 41, 40, 36) enjoy a significant edge. Steel (32), tobacco products (16), and beverages (15) exhibit low contributions to new product development.
- Across ownership classifications, R&D makes its largest contribution to new product sales in the foreign sector. Table 11 shows that in FOR enterprises, the marginal efficiency of R&D expenditure among firms reporting positive levels of R&D effort and new product sales is approximately 80 percent greater than that of the average SOE, which registers the lowest level of efficiency among the seven ownership groups. The higher measured returns in the foreign sector of new product sales with respect to R&D may exhibit a tendency for new products to be developed in the parent company outside of China and later deployed to the Chinese-based subsidiary with relatively little effort.

Performance. Tables 7 and 8 report the estimation results for the two performance equations that assess the impact of new product innovation of firm productivity and profitability. Because shocks to productivity (output) and profits are likely to be correlated with SALES used to define new product intensity (NPsales/SALES), for both the productivity and profit equations, we use, in the manner described in the previous section, instruments to correct for the potential downward bias in OLS estimates of the contribution of new product sales to productivity and profits.

Productivity. Table 7 shows the estimation results for the third equation in the model – the productivity equation. Controlling for inputs of capital, labor, and materials, estimates of equation (3) show that new product sales are associated with higher productivity. The estimated elasticity of output with respect to new product sales is 0.035. As shown in Table 11, this output elasticity translates into an 18-cent increase in output for every one-dollar increase in new product. One possible explanation for this seemingly small effect on output of new product sales is that the introduction of new products may result in the collateral retirement of old products. In addition, pursuant to quality improvements, the R&D reporting system may allow for the reclassification of the sales of existing products as "new product sales."

Table 11 shows that the marginal productivies of new products are comparatively high in the state-owned sector. This finding suggests that while SOEs appear to be inefficient in creating new knowledge, once they secure new knowledge, in the form of new products, they are able to use it effectively to achieve productivity gains. The comparatively small marginal productivities shown in the foreign (FOR) and overseas-

invested enterprises (HKT) may reflect high rates of turnover and retirement of existing products that they produce, such as footwear and apparel.

Profitability. The estimates reported in Table 8 show that new product sales contribute significantly to profitability. Our estimates show an average rate of return on new products of 10.6 percent. These estimated returns range at the high end from nearly 16.5 percent for the "other" ownership form and 14 percent for SOEs to an estimated 5.2 percent for overseas (HKT) enterprises at the low end. The comparatively low rates of return on new product innovation in the overseas sector may reflect the particularly intense competitive pressures in this sector, whereas the high returns for SOEs may reflect a relatively high degree of monopoly pricing in that sector. The pharmaceutical sector (27) stands out in the contribution of new product sales to industry profitability.

9. Reduced Form Estimation

In order to measure the full impact of R&D effort on performance, we substitute the knowledge production function, summarized by equation (3), into the productivity and profitability equations, equations (7) and (8). In the U.S. R&D literature, there is a well-established practice of investigating these reduced form functions, which examine the total effect of R&D input on firm performance. Like much of the production function literature, Griliches' (1964) first production function estimates with an R&D variable originated in an agricultural study. Griliches (1995) and the Australian Industry Commission (1995) provide comprehensive reviews of the more subsequent literature that built on Griliches' classic study.

To address the possibility of simultaneity bias arising from correlations between shocks to output and profit in 1999 and sales in 1998, we again use an IV for SALES. Arguably, in the presence of firm effects that persist overtime, R&D expenditure may also be endogenously driven by output and productivity. Later in this section, we examine the matter of possible endogeneity of R&D intensity.

Productivity. We estimate the reduced form productivity equation with and without the IVs and report the results in Table 9. Below we discuss the IV results.

- The positive and statistically significant estimate of the interactive term implies scale bias in the productivity of R&D expenditures. This result is confirmed by the estimates shown in Table 11, which show a somewhat higher marginal productivity for R&D expenditure in large-size enterprises as compared with their medium-size counterparts.
- As shown in Table 11, R&D expenditure in collectives and state-owned enterprises exhibit the highest marginal products, whereas the lowest marginal products of R&D are found in the overseas and foreign sectors. Again, depressed marginal products in the HKT and FOR sectors may reflect high rates of product turnover in these sectors.
- Table 9 shows that the Chinese industries in which R&D exhibits the greatest impact on productivity are tobacco, beverage, and the medical and pharmaceutical industries.

Profitability. Estimates for the restricted profit equation are shown in Table 10.[14,15] Overall, we find implausibly high returns to R&D in Chinese industry. As shown in Table 11, the elasticity of profit with respect to R&D is 0.125. We derive from this elasticity two measures of returns to R&D. One, based on all observations in our sample, is 277 percent; the second, based only on those firms for which profit and R&D expenditure are positive, is 156 percent. Below, we report additional results and then explain the source of upward bias and a method for recalibrating our initial calculations. Key results include the following:

- As with measures of productivity, Table 11 shows that returns to R&D expenditure tend to be higher in large-size enterprises (284%) as compared with their medium-size counterparts (247%).
- Apart from the OTHER sector, returns to R&D expenditure across ownership types do not exhibit dramatic differences. They range from a low return of 146 percent to a high of 404 percent for the other sector; for the state sector the measured return is 326 percent. Again, the low rate of return to innovation activity in the overseas sector may reflect the competitive nature of this sector.

We examine two possibilities that may explain the strikingly high estimates of returns to R&D in Chinese industry. One is the problem of endogeneity of R&D

[14] The profit function is restricted in the sense that we assume that the marginal product of intermediate inputs is flexibile, so that the marginal productivity of intermediate inputs (M) approximates zero. This assumption is consistent with the finding of Jefferson, Rawski, Wang, and Li (2000).

[15] As part of our investigation of the marginal profitability of R&D expenditure, we investigate the implication of having omitted the loss-making enterprises, which constitute five percent of our sample. We test the stability of profit function over the regression sample (5451 enterprises) and the sample of loss-making enterprises (289 enterprises). This exercise results in a conclusive rejection of the hypothesis of stability. Because of the asymmetric nature of relationship between R&D spending and profit, we do not attempt to implement a correction for selection bias resulting from the omission of the relatively small collection of loss-making enterprises.

intensity in equation (8). The other is the problem of having used an annual R&D expenditure measure to proxy the preferred R&D stock measure.

We are not optimistic about the prospect of creating an effective set of instruments for R&D intensity. In principle, we should use all of the predetermined variables in the model. The problem is that among the list of candidates, few are likely to be truly exogenous. Griliches (1998), speaks directly to this problem:

> …future output and its profitability depend on past R&D, while R&D, in turn, depends on both past output and the expectation about its future…In cross sectional data with only a few observations per firm, it is much harder to make (causality) distinctions, particularly since current expectations about the future are based on current and past data (p. 34).

Among the IV candidates are lagged values of sales, profit, and industry concentration ratio shown in the R&D effort equation (equation 2), the age variable in the knowledge production function, the capital stock and labor force in the performance and reduced form equations, and the ownership and industry classifications that appear in all of the estimation equations. We also have the size dummies that, along with the ownership and industry classifications, have been used to instrument sales. Among these, only these three instruments and age are clearly exogenous. Cohen and Levin (1989), for example, cite an extensive literature that investigates simultaneity between innovation and industry concentration (p. 1076). In addition, within the Solow model, capital accumulation results from productivity gain, which at the firm level is also likely to affect

labor demand. Indeed, our efforts to use various combinations of the above list of potential instruments for both R&D intensity and R&D alone in the restricted profit function in all cases led to elasticities of profit with respect to R&D that substantially exceed the estimate of 0.125 reported above and shown in Table 11. In fact, we find a relatively low correlation of coefficient in the association between lnPROF(99) and ln(RDE/SALES)(98) used in equation (8), i.e. 0.062 for the full sample of 5,451 observations and -0.119 for the subsample of observations for which profit and R&D are both positive. By comparison for R&D expenditure not in intensive form, we find correlation coefficients for lnPROF(99) and RDE(98) of 0.201 for the full sample and 0.488 for the sample with only positive observations. We draw two conclusions. First, with the available variables and just three years of data, we cannot resolve the endogeneity issue. Second, the correlation coefficients suggest that the potential for endogeneity is substantially less when we investigate the impact of inovation on firm performance using R&D intensity rather than using R&D alone. Moreover, we find evidence that if a suitable set of instruments were available, they might yield yet a higher estimate of the returns to R&D than that which we have reported.

The second candidate explanation for the strikingly high estimates of R&D returns shown in Table 11 is our use of a single year's R&D expenditure to measure R&D input rather than the appropriate measure of R&D stock. To investigate the sensitivity of our results to the construction of the R&D measure, we create a second measure, which is the sum of reported R&D expenditures for 1997 and 1998, i.e. [RDexp(97) + RDexp(98)]/sales(98). The estimation results are virtually identical to the original set of estimates used to compute the returns shown in Table 11. The estimated

elasticity of profit with respect to the two-year measure of R&D input rises to 0.127 from 0.125. We therefore assume that the relevant elasticity is stable over various flow and stock measures of R&D input.

For the purpose of constructing an R&D stock, we employ the following formula: RDstock(98) = [RDexp(98)/GFI(98)]*NVFA(98), where GFI is gross fixed investment in 1998, measured as the difference in gross fixed assets [i.e. OVFA(98) – OVFA (97)], and NVFA(98) is net value of fixed assets in 1998. This method of construction is founded on two assumptions: (i) the time structures of investment in fixed assets and R&D are similar, and (ii) the rates of depreciation for the two investment series are similar. Because within the population of LMEs, R&D investment is likely to be growing more rapidly than fixed investment and R&D assets are generally depreciated more rapidly than fixed assets, our estimate of RDstock(98) represents an upper-bound estimate.

For the enterprises reporting positive R&D expenditures in 1998, the ratio of PROF(98) to RDstock(98), used in conjunction with the relevant elasticity (0.127) to compute the marginal profit of R&D stock, is 2.850. The implied marginal return to the stock of R&D investment is 0.362, or about 23 percent of the comparable measure using the R&D expenditure flow (i.e. 1.563). As we see in Table 12, which shows the return to the stock of fixed investment being in the vicinity of 10 percent, our lower bound estimate of returns to R&D is approximately 3.6 times our estimate of the return to fixed investment.

Jaffe (1986) investigated the effects of R&D spending on profit in a sample of 432 U.S. manufacturing firms for 1973 and 1979. His estimates suggest a gross return to R&D expenditure of 27 percent, compared with just 15 percent for physical capital. With

relatively underdeveloped capital markets to fund R&D activities in China and a far smaller proportion of firms in China with established R&D operations, one might anticipate that the spread between returns to fixed capital and R&D would be greater than the factor of two found by Jaffe.

10. The channels of innovation

The three-equation model used in this paper is based on the assumption that innovation operates through the single channel of new product innovation. In fact, innovation may be broadly construed to span a much wider variety of activities than new product development. Other forms of innovation include improvements in product design and quality, as well as process innovation including the installation of new equipment and quality control methods. In a developing economy, such as China's, we should anticipate that much of the technical innovation spending, which we characterize as R&D, is unrelated to the development of new products. Indeed, the measure of technology innovation expenditure that we use in this paper, which is more broadly defined than the conventional measure of R&D, is intended to capture this more inclusive set of innovation activities.

In order to identify the proportion of returns to R&D effort that can be attributed to new products, we compute the contribution of R&D that operates through new product sales and compare this contribution with the total returns to R&D that we estimate from the reduced form equation. The result, shown in Table 13, indicates that 12 percent of the total return to R&D expenditure can be attributed to the production of new products.

This result is unchanged whether we use the full subsample of enterprises to compute returns or only those firms reporting positive observations. This estimate of 12 percent suggests that a substantial majority of the return to innovation in China's LMEs accrues from innovative activity that lies outside the province of new product development. We conclude that a substantial portion of the returns to R&D in Chinese industry results from improvements in existing products and forms of process innovation that are not captured by the new product innovation channel.

11. Conclusions

Using a panel of China's large and medium-size enterprises, we investigate patterns of R&D activity and its impact on firm-level performance. Because of missing or ambiguous data for a substantial portion of our population of China's large and medium-size industrial enterprises, our research draws inferences only for firms that report positive levels of activity in one or more areas of R&D and for which profit is non-negative. A key result of our analysis is that we find statistically significant relationships for the essential R&D relationships within the model. We identify several key findings.

First, R&D performers are more concentrated among capital-intensive firms with large sales volumes, particularly those classified by China's National Bureau of Statistics NBS as "large size". The highest incidence of R&D performers is located in equipment and machine manufacturing and in the pharmaceutical industry; the lowest concentrations are located in printing, garments, and timber. Across ownership types, controlling for size, capital intensity, and industry, R&D performers are more concentrated among SOEs

and shareholding companies and least concentrated among foreign and overseas enterprises.

Within our sample of R&D performers, which accounts for somewhat more than one-quarter of China's population of LMEs, we investigate the determinants of R&D intensity, the efficiency of new product development, and the effect of new product sales on firm performance. The return to new product sales accounts for approximately 12 percent of our estimate of the total return to R&D.

R&D intensity responds to firm size, profitability, and market concentration, although the robustness of these impacts varies with the inclusion of industry dummies and a lagged variable in R&D intensity. Within the context of a knowledge production function, we find a robust association between R&D intensity and new product sales. Our results show significant returns to new product sales, particularly in the pharmaceutical and telecommunications equipment sectors. State-owned enterprises exhibit the lowest efficiency in knowledge production. However, once they acquire new knowledge, SOEs appear to be able to use the innovations as effectively, or sometimes more so, than enterprises of other ownership forms.

Using a reduced form derived from the three-equation model including a restricted profit function, we find significant returns to R&D expenditure. The reduced form shows that the marginal profit of R&D expenditure is surprisingly high in state industry and lowest in the overseas (HKT) sector. By constructing a stock of R&D investment, we estimate that returns to investment in R&D are substantially higher than those to fixed investment – at least three to four times as high.

Table 1
Descriptive Statistics, 1999
(means) all manufacturing LMEs*

	Full sample	Large-size[1]	Medium-size[1]
Total number	19,943	6,846	13,097
Sales ('000 yuan)	191,655.3	416,995.6	63,666.2
NVFA[2] ('000 yuan)	159,488.8	361,755.1	44,605.2
Labor (persons)	1,469.5	2,859.1	680.3
Profit/sales	10.34%	11.25%	9.83%

[1]Various measures of sales, capacity, and value of fixed assets are used by the National Bureau of Statistics to distinguish between "large size" and "medium size" industrial enterprises. These measures are set forth in the document "Criteria for defining large- and medium-size enterprises," National Bureau of Statistics (2000), Beijing.
[2]NVFA = net value of fixed assets, measured as the sum of the depreciated value of annual investment measured in terms of prices that were current for the year of the investment.

Table 2
R&D Statistics, 1995 and 1999
(means) all manufacturing LMEs[1]

Variable	Full population		Observations > 0	
year	1995	1999	1995	1999
R&D expenditure/ Sales (%)	0.4	0.8	2.1 (3,821)[1]	3.3 (5,050)
R&D personnel/ total employment (%)	3.4	7.3	5.4 (12,084)	16.5 (8,708)
New product sales/ total sales (%)	8.8	11.8	19.4 (8,551)	33.1 (6,168)
Patent applications (average/firm)	0.175	0.362	2.86 (1,157)	5.43 (1,168)
No. of enterprises	20,781	19,943	12,235	14,037[2]

[1]The figures in parentheses represent the number of enterprises, which report positive levels of activity for the relevant R&D measure.
[2] Omits observations for which RDexp = RDperson = NP (new products) = PATAP (patent applications) = 0 and profit are all zero. This subset of firms is the sub-sample used for the probit estimates reported in Table 3.

Table 3
Probit estimates of marginal probabilities of R&D performance

Variable/enterprise type	Marginal probability of inclusion[1]	Variable/enterprise type	Marginal probability of inclusion[1]
Large	0.105*	34	0.138*
Sales	0.105*	26	0.137*
COE	-0.120*	33	0.060*
FOR	-0.368*	28	0.055
HKT	-0.242*	15	0.050*
STK	-0.128	17	0.025
PRI	-0.226*	14	0.020
OTH	-0.120*	43	0.012
By industry:	-	19	0.006
42	0.393*	16	0.002
36	0.348*	25	-0.012
35	0.344*	32	-0.022
27	0.339*	22	-0.026
41	0.290*	20	-0.075
40	0.283*	18	-0.115*
37	0.253*	23	-0.199*
29	0.231*	pred. P (at x-bar)	0.519
24	0.202*	Pseudo R^2	0.178
21	0.183*	No. of obs.	19,165
30	0.148*		

[1] Reference = medium-size, SOE, industry 31.
* Significant at the 5% level of statistical significance.

Table 4
Distribution of R&D Performers, balanced sample (1997-99)

	Full sample	med	large	SOE	COE	FOR	HKT	STK	PRI	OTH
Full sample distribution, 1999	19,943	13,097	6,846	9,354	3,350	1,937	1,517	3,283	313	189
Regression sample[1]	5,451	2,881	2,570	3,148	803	279	262	643	261	55
Of which:										
(1) RDexp98>0	3,052	1,424	1,628	1,833	356	170	150	321	192	30
(2) RDper98>0	4,957	2,519	2,438	2,958	660	250	233	222	585	49
(3) NPsales>0	3,710	1,783	1,923	2,188	487	208	182	373	228	44
(4) PATapp>0	761	279	482	429	96	38	38	106	51	3
(5) Profit ≥ 0	5,451	2,881	2,570	3,148	803	279	262	643	261	55

[1]The regression sample used in this paper consists of enterprises that report at least one positive value for (1) to (4) and also non-negative profit (5).

Table 5
Determinants of R&D intensity [equation (2)]
dependent variable = ln[(RDexp/SALES)(98)]

Variable	Without lag of the dep. variable		With lag of the dep. var. (α_5)
	(1)	(2)	(3)
Constant	-3.280 (6.903)	-20.928 (0.577)	-22.243 (0.637)
lnSALES(97)(IV)	-0.201 (6.385)	1.466 (0.440)	1.633 (0.510)
lnPROF(97)	-0.077 (5.755)	-0.091 (6.784)	0.026 (1.861)
lnCR2(98)	0.614 (3.023)	0.301 (1.390)	0.109 (0.525)
lnCR2(98)*sq	-0.082 (2.573)	-0.040 (1.177)	-0.014 (0.433)
ln[RDexp/ SALES(97)]	n.a.	n.a.	0.265 (20.821)
DlnRDexp(98) [2]	-4.010 (116.477)	-3.987 (115.764)	-3.913 (11.724)
DPROF(97) [2]	-2.019 (1.594)	-1.915 (1.528)	0.524 (0.432)
DlnRDexp(97) [2]	n.a.	n.a.	4.024 (20.045)
OWN[3]	(+) FOR,HKT (-) COE	-	-
IND[3]	NO	(-) 34	-
Adj R-square	0.718	0.726	0.746
Obs.	5451	5451	5451

[1] Figures in parentheses in Tables 5 through Table 9 are t-statistics.
[2] Values of PROF(97), RDexp(98), and RDexp(97) = 0 have been converted to 0.0001k and represented by dummy variables. See Section 7 and FN # 11.
[3] Estimates of reported ownership and industry classifications are significant at ≥ 95 level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 6
Knowledge production function [equation 3]
dependent variable = ln[NP/SALES(99)]

Variables	OLS	IV[1]
Constant	-1.694 (15.030)	-1.696 (15.030)
lnAGE	-0.016 (0.965)	-0.016 (0.989)
ln[RDexp/ SALES](98)	0.191 (6.766)	0.202 (7.436)
ln(RDexp/SALES) *ln(SALES)(98)	-0.008 (3.805)	-0.009 (4.801)
Dln(NP/SALES)[2]	-6.859 (182.474)	-6.854 (183.821)
Dln(RDexp)[2]	0.397 (6.020)	1.513 (8.138)
OWN[3]	(+) FOR	
IND[3]	(+) 41, 40, 36 (-) 32, 16, 15	
Adj. R-square	0.878	0.878
Obs.	5451	5451

[1]The instrumental variables are size [i.e. Large (2 discrete sizes) or Medium (3 discrete sizes)], ownership (OWN), and industry classification (IND).
[2]Values of NP/SALES = 0 and RDexp = 0 have been converted to 0.0001k and represented by dummy variables. See Section 7 and FN # 11.
[3] Estimates of reported ownership and industry classifications are significant at ≥ the 95% level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 7
Impact of R&D output [ln(NP/SALES)] on productivity [equation 7]
dependent variable = lnQ(99)

Variable	OLS	IV[1]
Constant	0.881 (24.446)	1.047 (28.842)
lnK (capital)	0.120 (22.840)	0.112 (27.489)
lnL (labor)	0.121 (29.228)	0.116 (22.679)
lnM (materials)[2]	0.749 (n.a.)	0.749 (n.a.)
ln(NP/SALES)(99)	0.013 (5.091)	0.035 (17.473)
Dln(NP/SALES)(99)[3]	0.038 (1.968)	0.578 (15.711
OWN	(+) FOR, HKT, COE, STK	
IND[4]	(+) 40, 15, 26, 14, 11, 12, 24, 37, 22, 25, 31, 33, 19, 21, 36, 17, 30	
Adj R-square	0.970	0.971
Obs.	5451	5451

[1]The instrumental variables are size [i.e. Large (2 discrete sizes) or Medium (3 discrete sizes)], ownership (OWN), and industry classification (IND).
[2]Restricted to equal the value share of intermediate inputs in total output as computed within the sample.
[3]Values of NP/SALES(99) have been converted to 0.0001k and represented by a dummy variable. See Section 7 and FN # 11.
[4]Estimates of reported ownership and industry classifications are significant at ≥ the 95% level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 8
Impact of R&D output (new product sales) on profit* [equation 8]
dependent variable = lnPROF(99)

Variable	OLS	IV[1]
Constant	0.026 (0.169)	0.744 (4.881)
lnK (capital)	0.549 (31.605)	0.510 (29.886)
lnL (labor)	0.454 (20.586)	0.437 (20.379)
ln(NP/SALES)(99)	0.056 (5.140)	0.149 (17.917)
D[PROF(99)]	-17.240 (48.252)	-17.244 (49.559)
D[NP(99)][2]	0.049 (0.610)	2.380 (15.408)
OWN	(+) FOR, HKT, COE, STK, PRI	
IND[3]	(+) 27,16,25,14,26,23	
Adj. R-square	0.496	0.666
Observations	5451	5451

[1]The instrumental variables are size [i.e. Large (2 discrete sizes) or Medium (3 discrete sizes)], ownership (OWN), and industry classification (IND).

[2]Values of PROF(99) and NP/SALES (99) = 0 have been converted to 0.0001k and represented by dummy variables. See Section 7 and FN # 11.

[2] Estimates of reported ownership and industry classifications are significant at ≥ the 95% level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 9
Reduced Form: Total effect of R&D effort on productivity
Dependent variable = lnQ(99)

Variable	R&D expenditure	
	OLS	IV[1]
Constant	2.149 (48.718)	0.909 (24.47)
lnK (capital)	0.049 (12.500)	0.116 (27.54)
lnL (labor)	0.052 (11.001)	0.115 (21.85)
lnM (materials)[2]	0.749 (n.a.)	0.749 (n.a.)
ln[RDexp/ SALES(98)]	0.241 (40.108)	-0.045 (5.65)
ln[RDexp/SALE) *ln[SALES(98)]	-0.022 (44.835)	0.005 (6.51)
D[RDexp(98)][3]	-0.007 (0.647)	-0.008 (1.15)
OWN[4]	(+) FOR, COE, HKT, STK	
IND[4]	(+) 16, 15, 27, 23, 25, 41, 17, 18, 40	
Adj. R-square	0.969	
Obs.	5451	5451

[1]The instrumental variables for SALE are size [i.e. large (3 discrete sizes) or medium (2 discrete sizes)], ownership (OWN), and industry classification (IND).
[2]Restricted to equal the value share of intermediate inputs in total output as computed within the sample.
[3]Values of RDexp = 0 have been converted to 0.0001k and represented by a dummy variable. See Section 7 and FN # 11.
[4] Estimates of reported ownership and industry classifications are significant at ≥ the 95% level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 10
Reduced Form: Total effect of R&D effort on profitability
Dependent variable = lnPROF(99)

Variable	RDexp(98) only	
	OLS	IV[1]
Constant	5.412 (29.111)	0.852 (4.597)
lnK (capital)	0.244 (14.827)	0.521 (29.205)
lnL (labor)	0.173 (8.656)	0.442 (19.919)
ln[RDexp/ SALES(98)]	1.059 (41.769)	0.250 (10.299)
ln[RDexp/SALES) *lnSALES(98)]	-0.093 (45.715)	-0.011 (6.700)
D[PROF(99)] [2]	-17.286 (56.190)	-17.393 (48.740)
D[RDexp(98)] [2]	0.094 (2.056)	1.827 (11.948)
OWN[3]	(+) FOR, COE, HKT, STK, PRI	
IND[3]	(+) 27, 41, 16, 15, 23, 25, 26	
Adj. R-square	0.739	0.649
Obs.	5,451	5,451

[1]The instrumental variables for SALES are size [i.e. Large (3 discrete sizes) or Medium (2 discrete sizes)], ownership (OWN), and industry classification (IND).
[2]Values of PROF(98) = RDexp(98) = 0 have been converted to 0.0001k and represented by a dummy variable. See Section 7 and FN # 11.
[3] Estimates of reported ownership and industry classifications are significant at ≥ the 95% level of statistical significance. Ownership and industries are ordered from the most to the least statistically significant.

Table 11 (all using IV)
Estimates of elasticities and marginal products

	Elasticity	Full Sample	Med	Large	SOE	COE	FOR	HKT	STK	PRI	OTH
dNP(99)/dRD(98) (from Table 6)											
1. Including all observations	0.101	3.153	2.905	3.199	2.784	3.365	5.249	3.335	3.032	3.397	2.942
2. Omiting zero obs. from both numerator and denominator	0.101	2.710	2.465	2.784	2.384	2.910	4.312	3.181	2.556	2.960	2.102
dQ(99)/dNP(99) (from Table 7)											
1. Including all observations	0.035	0.183	0.193	0.182	0.247	0.245	0.086	0.083	0.168	0.152	0.301
2. Omiting zero obs. from both numerator and denominator	0.035	0.120	0.116	0.131	0.164	0.144	0.060	0.055	0.113	0.104	0.230
dProf(99)/dNP(99) (from Table 8)											
1. Including all observations	0.035	0.106	0.102	0.106	0.140	0.113	0.060	0.052	0.099	0.088	0.165
2. Omiting zero obs. from both numerator and denominator	0.035	0.070	0.061	0.077	0.094	0.067	0.042	0.035	0.066	0.060	0.126
dQ(99)/dRD(98) (from Table 9)											
1. Including all observations	0.007	1.163	1.126	1.169	1.382	1.656	0.909	0.554	1.025	1.042	1.781
2. Omiting zero obs. from both numerator and denominator	0.007	0.656	0.573	0.732	0.789	0.845	0.524	0.351	0.582	0.621	0.971
dProf(99)/dRD(98) (from Table 10)											
1. Including all observations	0.125	2.770	2.472	2.838	3.255	3.185	2.616	1.456	2.496	2.502	4.041
2. Omiting zero obs. from both numerator and denominator	0.125	1.563	1.262	1.778	1.862	1.628	1.509	0.923	1.418	1.490	2.204

Table 12
Marginal productivities based on the restricted profit function[1]

Measure	Elasticities	Marginal profit (yuan)
dPROF/dLAB	0.442	9,645
dPROF/dNVFA	0.521	0.109
dPROF/dRDexp(98)	0.125	2.770/ 1.550[2]
dPROF/dRDStock(98)	0.127	0.362

[1]derived from Table 11.
[2]The figure above the / is the mean of PROF/RD for the full sample; the figure below the / is computed for observations with positive values of PROF and RDexp.

Table 13
Channels of R&D impact[1]

(A) dPROF/ dRDexp	(B) ∂PROF/∂NP x ∂NP/∂RDexp	(B)/(A)
2.770	(0.106 x 3.153) = 0.334	12%
1.571	(0.070 x 2.710) = 0.190	12%

[1]derived from Table 11.

References

Australian Industry Commission, 1995, "Research and Development," Report N. 44, Government Publishing Service, Canberra.

Arrow, Kenneth. 1962, "Economic Welfare and the Allocation of Resources for Invention,"in *The Rate and Direction of Inventive Activity*, NBER, Princeton University Press.

Basant, R. and B. Fikkert. 1996, "The Effects of R&D, foreign Technology Purchase, and International and Domestic Spillovers on Productivity in Indian Firms," *The Review of Economics and Statistics*, 78,2: 187-199,

Cohen, W.M, R.C. Levin, D.C. Mowrey, 1985, "Firm Size and R and D Intensity: A Re-examination, *Journal of Industrial Economics*, 35:543-65.

Cohen, W.M. and Klepper, S. 1996, "A Reprise of Size and R&D," *The Economic Journal,* 106:925-951.

Cohen, W.M. and Levin, R.C. 1989, "Empirical Studies of Innovation and Market Structure," in R. Schmalansee and R.D. Willig, eds., *Handbook of Industrial Organization*, Amsterdam: North Holland: 1059-1107.

Comanor, W.S. and Scherer, F.M. 1969, "Patent Statistics as a Measure of Technical Change," *Journal of Political Economy*, 77:392-8.

Crepon, , Bruno, Emmanuel Duguet, and Jacques Mairesse. 1998, "Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level," NBER Working Paper Series, working paper 6696.

Grabowski, 1968. "The Determinants of R and D in Three Industries," *Journal of Political Economy*, 76:292-306.

Griliches, Zvi. 1964, "Research Expenditures, Education, and the Aggregate Production Function," *American Economic Review*, 54(6):961-74.

Griliches, Zvi. 1995, "R&D and Productivity: Econometric Results and Measurement Issues," in *Handbook of the Economics of Innovation and Technical Change*, ed. P. Stoneman, Basil Blackwell, Oxford.

Griliches, Zvi, and J. Mairesse. 1984, *R and D, Patents and Productivity*, NBER, University of Chicago Press.

Griliches, Zvi.1998, *R&D and Productivity*, University of Chicago Press, Chicago.

Hay, Donald A. and Derek J. Morris. 1991, *Industrial Economics and Organization* (second edition), Oxford University Press.

Hu, Albert G.Z. May 1999, "Institutions, Organizations, and Technological Innovation in Chinese Industry" (dissertation) Graduate School of International Economics and Finance, Brandeis University.

Jaffe, Adam B. 1986, "Technological Opportunity and Spillovers of R and D: Evidence form Firms' Patents, Profits and Market Value, *American Economic Review*, 76:984-99.

Jefferson, Gary H., David H.M.Bai, Guan Xiaojing , and Yu Xiaoyun. (forthcoming), "Ownership, Performance, and Innovation in China's Large and Medium-Size Industrial Enterprise Sector," *China Economic Review*.

Jefferson, Gary H.,. Rawski, Thomas G, Wang Li, and Zheng Yuxin. 200, ""Ownership, Productivity Change, and Financial Performance in Chinese Industry," *Journal of Comparative Economics*, 28, 786-813.

Levin, Richard C., W.M. Cohen, and D.C. Mowrey, 1975. "R and D Appropriability, Opportunity and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review*, 75, 20-24.

Loof, Hans (hans.loof@lector.kth.se) and Almas Heshmati. 2001, "On the Relationship Between Innovation and Performance: A Sensitivity Analysis," unpublished manuscript.

Schankerman, Mark. 1981, "The Effects of Double-Counting and Expensing on the Measured Returns to R&D," *Review of Economics and Statistics*, 63,3:454-458.

Scherer, F.M. 1967, "Market Structure and the Employment of Scientists and Engineers," *American Economic Review*, 57:524-531.

Schumpeter, Joseph. 1942, *Capitalism, Socialism, and Democracy*, Harper and Brothers, N.Y.

Scott, J.T. 1984, "Firm Versus Industry Variability in R&D Intensity," in Z. Griliches, ed., *R&D, Patents, and Productivity*, University of Chicago Press, Chicago.

Van Leeuwen, George (glwn@cbs.nl) and Luuk Klomp. 2001, "On the Contribution of Innovation to Multi-factor Productivity Growth," unpublished manuscript.

Annex 1
Ownership Code

SOE – state-owned enterprise
COE – collective-owned enterprise [includes township and village-owned enterprise (TVE)]
FOE (foreign) – a firm that includes foreign-owned assets
HKT (overseas) – a firm that includes assets from investors in Hong Kong, Macao, or Taiwan
STK – a shareholding or joint stock firm
PRI – a privately owned firm
OTH – an enterprise that is not classified as any of the above

Annex 2
Industry Code

13 Food processing
14 Food production
15 Beverage production
16 Tobacco processing
17 Textile industry
18 Garments and other fiber products
19 Leather, furs, down and related products
20 Timber, bamboo, cane, palm fiber and straw products
21 Furniture manufacturing
22 Papermaking and paper products
23 Printing and record medium reproduction
24 Cultural, educational and sports goods
25 Petroleum processing and coking
26 Raw chemical materials and chemical products
27 Medical and pharmaceutical products
28 Chemical fiber
29 Rubber products
30 Plastic products
31 Nonmetal mineral products
32 Smelting and pressing of ferrous metals
33 Smelting and pressing of nonferrous metals
34 Metal products
35 Ordinary machinery manufacturing
36 Special purpose equipment manufacturing
37 Transport equipment manufacturing
40 Electric equipment and machinery
41 Electronic and telecommunications equipment
42 Instruments, meters, cultural and office machinery
43 Other manufacturing