

The Role of Research and Innovation in Promoting Productivity in Chile^{*}

Jose Miguel Benavente⁺

Third Draft
November 4, 2002

Abstract

This paper continues the empirical research line started by Crepon et al. (1998) about the impact of research and development on innovation and innovation on productivity of firms. In this paper we estimate a structural model using Asymptotic Least Squares (ALS) which corrects for selectivity and simultaneity biases taking in consideration the particular characteristics of the available data. We find that most of the Schumpeterian hypotheses are confirmed: research and innovative activities are related with firm size and market power. However, in the case of Chile, firms productivity is not affected by innovative results nor by research expenditures in the short run.

Keywords : R&D, innovation, productivity, Asymptotic Least Squares, structural estimation and censored data.

JEL classification : C13, O32.

Correspondence : jbenaven@econ.uchile.cl; Department of Economics, University of Chile, Diagonal Paraguay 257 of. 1503, Santiago, Chile.

^{*} This paper is based on Chapter 6 of my DPhil thesis at Oxford. I wish to thank Bronwyn Hall, Paul David, Nick von Tunzelmann, Jacques Mairesse, Emmanuel Duguet, Valpy FitzGerald, participants at the Gorman workshop at Oxford, and two anonymous referees for their comments and suggestions.

⁺ Department of Economics, University of Chile.

1 Introduction

This paper continues the empirical research line started by Crepon et al (1998) about the impact of research and development on innovation and innovation on productivity of firms. It is based on a model which takes into account the whole process of innovation that includes decisions of firms to engage in R&D activities, the results of these efforts and their impact on productivity. In this paper we study a less developed country case: Chile, for which we have data about innovation, research and development and production activities and its results during the period 1995-1998.

This is the first attempt to analyse research and innovation activities in a LDC and its further impacts on productivity. As in previous studies this study presents two main characteristics. On the one hand, it focuses both on innovation input (e.g. R&D) as well as on innovative output like innovative sales. That is, firms invest in research and development aiming either to launch new products and/or reduce their production costs. These, if successfully accepted by the market, could have an impact on the firms production performance and economic viability. The model takes into account the allocation of resources to research and development activities, the results of the innovation process (where R&D expenditure is one of its determinants), and productivity, where together with capital and labour, innovation performance is also included as one of its determinants.

On the other hand, and given the nature of the data: censored, by interval, truncated together with some already known problems of selectivity and simultaneity, we have followed the steps of Crepon et al (1998) by adopting a new technique on the analysis of the data. We rely on generalised tobit estimation to deal with problems of selectivity and on an ordered probit in the case of interval data all this implemented in an two stage estimation procedure to account for simultaneity biases in the system. The first stage consist in a method of moments estimation while the second stage uses the asymptotic least squares to obtain consistent estimates of the structural parameters.

The assumptions made for the estimation are rather reasonable, and we have tested some of them using the latest available techniques. However, the main drawbacks of this study are first the cross-sectional nature of the

data and second that one of the dependent variables, R&D expenditures, is measured as a flow variable.

The paper is organised as follows. We start by presenting the basic statistics of the Chilean case together with the main hypotheses that will be tested. The presentation of the empirical framework, the equations and variables is the subject of section 3. In section 4 we present the results of the estimation and its discussion. Concluding remarks and suggestions for further research is the subject of section 5. An appendix gives details of the construction of the data, main sources and definition of the sample following by the comparison of our results with those obtained by using traditional econometric techniques.

2 Basic Statistics and Main Hypotheses

We start by presenting the main characteristics of firms that we have surveyed. We report results for the sample of firms surveyed and extrapolate these results to the whole universe of manufacturing firms using sample weights (inflation factors).

First some comments on the research and innovative variables. According to the Chilean Innovation Survey¹ about 45% of the firms claim that they spend resources on R&D activities. However, once the inflation factors are applied only about 15% of the total are engaged in R&D activities. This latter figure seems more realistic for a LDC. This significant difference is explained by the fact that the sample was designed in a way that firms which represent either main actors in their markets or are considerably larger (in terms of number of workers) compared with the rest of their competitors, have an inflation factor of one. This can be clearly observed in the average number of workers in the innovation sample, and also in terms of market share, where firms in this sample have almost three times more market power and size than the average firm in the universe of plants.

¹ See Benavente & Crespi (1999) for a detailed analysis of this survey.

[Table 1 about here]

When firms were asked about the introduction of a new product and/or a new process into the market, 30% of them claim to have done this in the last three years. But once extrapolated to the whole universe of firms in Chile this figure is reduced to 18%.²

In terms of the share of innovative sales in all the sales³, Chilean results show that they are heavily skewed. Almost half of the firms declare that innovative sales represents less than 10% of the total sales. Also, only 20% of the plants answered that innovative sales represent more than 30% of their sales during 1998. On the other hand, when managers were asked about the importance of demand-pull and technology push factors in the innovation dynamics of the firms, it appears that both factors are of considerable importance in the innovation dynamics of Chilean firms. Almost two third of the managers surveyed consider that these factors affect, from moderately to strongly, their propensity to innovate, where technology push indicators are much stronger than demand pull elements. This result is to be expected in a LDC where important channels of diffusion of technology are through FDI, purchasing of new machinery and the introduction of new inputs usually developed overseas and where demand elements, like customer demands, are less relevant.

Given the available data we aim to test most of the traditional hypotheses related with industrial innovation.⁴ These studies stressed the favourable position that for innovation could have for large firms with a good degree of market share by exploiting not only their size but also their market power.

On the one hand, market concentration allows the existence of monopolistic rents that enable firms within the industry to finance R&D

² A similar study of the French industry (Crepon et al. 1998) shows that on average, they have roughly the same market share but are less diversified than their Chilean counterparts. In terms of size the Chileans are slightly bigger and, surprisingly, there exist more firms in Chile that report R&D activities (around 15%) than in France (10%).

³ As is explained in Appendix I, the average market share and the diversification index were obtained from the firms sales desegregated by product lines as given by the Chilean Industry Survey (ENIA).

⁴ For a survey of them see Cohen & Levin (1989), Cohen (1995), and Cohen and Klepper (1996).

projects. In a perfectly competitive world private investment in R&D would not be feasible. Additionally, when imperfections in capital markets are recognised, large firms tend to have greater capabilities of securing the necessary resources to finance R&D projects bearing uninsurable risk. Moreover, the potential total impact of the results of an R&D project can be significantly greater in large firms given larger sale volumes. Hence, a product or process innovation that allows for an increase in the price-cost margin, will have greater absolute effects on profits on a firm with greater sales. Finally, larger plants have greater absolute incentives to improve internal process technologies, which in conjunction with economies of scale should lead firms to make greater relative efforts on process innovations than their smaller counterparts.

However, there exist several external forces which stimulate firms to innovate despite their size and market power. The first of such forces is based on demand factors, such as market growth. Schmookler (1966) first formulated such rationale known as the “demand-pull hypothesis”. On the other side, the role of scientific advancements in stimulating industrial innovative efforts may influence the path and rate of technology advance. The rationale was that advances in science enabled “technology-push” based innovations⁵ through the development of new concepts or when incorporated in new machinery and/or inputs. The relative opportunities to innovate within a given industry based on scientific progress form the basis for the notion of technology opportunities.

The empirical estimation procedure aims to capture the effect of some of these variables over the research and innovation effort of firms. It consists in four main equations: two for research and development, one for innovation and one for productivity. Given the design of the Chilean Innovation Survey some restrictions are imposed on the data which requires a different econometric treatment. We will estimate a reduced version of the model which includes only firm characteristics such as size and market share, and an expanded version including demand pull and technological push variables.

⁵ Firstly mentioned by Rosemberg (1974).

3 Empirical Model and Strategy for Estimation

In this section we empirically model the whole innovation process. We assume that there exists a set of firm characteristics, market structure and technological variables which conditions and shapes investments in R&D as well as the successful introduction of new products and processes into markets.

As in Crepon et al. (1998) we model the process of innovation with four main equations, two for research and development, one for innovation and one for productivity. It is important to note that each equation requires a different econometric treatment depending on the characteristics of the data. To make this work comparable to the French case, we will estimate a reduced version of the model which includes only firm characteristics such as size and market share, and an expanded version including demand pull and technological push variables. The definition of variables and the econometric specification of each equation are the subject of the next sections.

3.1 Research Equations

We will rely on a Generalised Tobit framework to model the decision to invest and the amount invested in research activities. It consists of two separate equations: one related to the decision to invest in R&D and the other to the amount of resources involved.

More precisely, we assume that there exists a latent dependent variable g_i^* for the firm i given by the following equation :

$$g_i^* = x_{i0} b_0 + u_{i0} \tag{1}$$

where x_{i0} is a vector of explanatory variables, b_0 the associated coefficient vector and u_{i0} an error term, where g_i^* represent some decision criterion like the expected present benefits associated with a research project. The econometrician observes that resources are invested in R&D activities if g_i^* is positive or bigger than a given threshold.

We then assume that a latent or true intensity of research k_i^* for firm i is determined by a second equation :

$$k_i^* = x_{i1}b_1 + u_{i1} \quad (2)$$

where $k_i^* = k_i$ the current R&D expenditure of firm i when it is engaged in research activities i.e. when g_i^* is bigger than the minimum industrial threshold. Here k_i^* and k_i are both expressed in logarithms, x_{i1} is a vector of the explanatory variables, b_1 the associated coefficient vector and u_{i1} an error term that summarises omitted determinants and other sources of unobserved heterogeneity.

Given that k_i^* is only observable when g_i^* is bigger than the minimum threshold, we need to specify their joint distribution in order to obtain an estimable model. Following the standard literature about generalised Tobit models we assume that the error terms in equation (1) and (2) follow a joint normal distribution :

$$\begin{pmatrix} u_{i0} \\ u_{i1} \end{pmatrix} \xrightarrow{iid} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right) \quad (3)$$

where σ_0 and σ_1 are the standard errors of u_{i0} and u_{i1} respectively and ρ is their correlation coefficient.

In the implementation of the model we use a flow variable for research expressed in thousand of Chilean pesos per worker given our available data. We are particularly interested to test the impact of firm size and market power on the probability as well as on the amount of resources invested in research activities. More precisely,

$$x_{i0} = x_{i1} = (l_i, s_i^w, d_i, \delta_i^1, \delta_i^2, \delta_i^3, \tau_i^1, \tau_i^2, \tau_i^3, S_i^1, \dots, S_i^9)$$

where l_i is employment, s_i^w the one year lagged average market share, and d_i the equivalent number of industry segments, all three variables expressed

in logarithms. As in previous works, δ_k and τ_k with $k = 1, 2, 3$ are two sets of demand-pull and technological push dummies. Finally, the S_i^j are nine industry dummies equal to one if firm i belongs to industry j and zero otherwise. These later dummies replace the constant term so that each industry has a different intercept. This is consistent with the design of the survey, which contains strata for different productive sector up to 2 digits ISIC (i.e., nine sectors).

As suggested in the last section we expect a positive impact of size on investments in R&D once controlled for industry characteristics. On the other hand, a positive relationship between market power and the probability of reporting R&D expenditure is expected. We also expect an impact from technological push and demand pull variables on research investments once controlled for other characteristics of the industries. In this later case, we are particularly interested in the significance of the impact rather than the sign since given the design of the variables negative impact is unfeasible.

3.2 The Innovation Equations

In general, it is assumed that innovative output is related to enhancements in the productivity capacity of the firm. However, in empirical work there exist several ways to proxy innovation output. More common proxies are the number of patents or share of innovative sales. The Chilean Innovation Survey only provides information on the later since patents are very scarce. The firm is asked what is the percentage share of its 1998 sales coming from products launched in the last three years 1996-1998 and/or what are the result of the implementation of absolutely new production techniques during the same period. The firm has to answer on a four-point scale : 0-10%, 10-30%, 30-70% and 70-100%.

We will assume that the underlying (unobserved) true share of sales t_i^* can be viewed as an innovation intensity variable where innovation is measured by the number of sales weighted innovations. Since the share is only known by intervals, we can specify the innovative sales equation as an Ordered Probit model :

$$t_i^* = \alpha_k k_i^* + x_{i2} b_2 + u_{i2} \quad (4)$$

where t_i^* is the underlying (unobservable) true share expressed as a logarithm, x_{i2} is a vector of explanatory variables, b_2 the associated coefficient vector and u_{i2} an error term which we will assume to be normally distributed with mean zero and variance σ_2^2 . The coefficient α_k is the elasticity of the percentage share of innovative sales relative to research expenditure, a measure of the return or impact of research on innovation output, x_2 is a vector of exogenous variables :

$$x_{i2} = (l_i, \delta_i^1, \delta_i^2, \delta_i^3, \tau_i^1, \tau_i^2, \tau_i^3, S_i^1, \dots, S_i^9)$$

with the same notation as above. Note that it is assumed that market share and diversification do not enter directly in the innovation equation but only indirectly through research in order to identify the system. By contrast, we assume that demand-pull and technological push factors could affect innovation output directly and indirectly. Also, including size in the innovation equation allows us to test whether the effect of firm size on innovation passes completely through the size of research activities.

As in the research equations, we include sectorial dummies which captures technological differences between markets not captured by technology-push or demand-pull variables.

3.3 The productivity equation

The last equation of the system is that for labour productivity. Following the standard literature we assume a Cobb-Douglas production function with physical capital and employment augmented with skill composition and innovation output. Then we have :

$$q_i = \alpha_l t_i^* + x_{i3} b_3 + u_{i3} \quad (5)$$

where q_i stands for labour productivity measured as a logarithm. The vector of the factors, other than innovation output, is :

$$x_{i3} = (l_i, c_i, E_i, A_i, S_i^1, \dots, S_i^9)$$

with c_i being the physical capital per employee as a logarithm, and E_i and A_i being the share of engineers and administrators respectively in the total number of employees. The coefficient α_i is the elasticity of total factor productivity with respect to innovation output, and b_3 is a vector of coefficients related to the elasticity of scale, that of physical capital and the skill composition parameters, reflecting percentage differences in efficiency of skilled labour relative to unskilled labour respectively.⁶

3.4 The overall model and its estimation

All four equations form a recursive non-linear system where innovation output is proxied by the ratio of innovative sales over total sales. The first two equations concern the R&D decision and expenditure :

$$g_i^* = x_{i0} b_0 + u_{i0} \tag{1}$$

$$k_i^* = x_{i1} b_1 + u_{i1} \tag{2}$$

⁶ Following Crepon et al. (1998) the interpretation of the parameters of skill composition is the following. Assume that labour, corrected for quality, enters the production function as: $L^* = f_U L_U + f_E L_E + f_A L_A$ instead of the uncorrected total employment $L = L_U + L_E + L_A$ where U stands for unskilled, E for engineers and A for administrators. Then the first expression can be re-written as: $L^* = f_U (L - L_E - L_A) + f_E L_E + f_A L_A$ or as $L^* = f_U L \{1 + (f_E / f_U - 1) L_E / L + (f_A / f_U - 1) L_A / L\}$. Using logarithm on both sides results in $\ln L^* \cong \ln f_U + \ln L + u_E E + u_A A$ and thus the coefficients of E and A in the production function are respectively: $b_{3E} \approx (\partial q / \partial l) u_E$ and $b_{3A} \approx (\partial q / \partial l) u_A$.

while the two last equations, for innovation output and productivity, are the following :

$$t_i^* = \alpha_1 k_i^* + x_{i2} b_2 + u_{i2} \quad (4)$$

$$q_i = \alpha_k t_i^* + x_{i3} b_3 + u_{i3} \quad (5)$$

No restrictions are imposed on the correlations between the disturbances u_{i0} , u_{i1} , u_{i2} and u_{i3} . However, we want to take into account the problems of selectivity and simultaneity that can arise due to the nature of the data (truncated, censored, discrete and interval) and the endogeneity structure of some of the equations (research expenditure is endogenous in the innovation equation and innovation is also endogenous in the productivity equation).

There are several ways to estimate this system of equations. As suggested by Crepon et al. (1998), Maximum likelihood would be impractical due to the non-closed form of the joint distribution. Based on studies of Lee (1981), they favour the use of the Asymptotic Least Squares estimator (ALS) also known as the Minimum Distance estimator (MD). This estimator is relatively more efficient and its computational costs lower compared with alternative estimators like the GMM.

4 Results

4.1 Results of the Basic model

Table 2 presents the results using ALS for the basic model. Results obtained with more traditional econometric techniques like OLS, 2SLS, ML and ALS for the equations estimated separately, are presented and discussed in the Appendix II.

The first two columns of Table 2 show the estimates of the research equations and the last two the estimates for the innovative sales equation and the productivity equation with innovative sales. The bottom line shows

the results of the Jarque Bera and Lee (1981) for normality in the probit model and the Pagan and Vella (1989) test for normality in the tobit model.

Starting with the decision to allocate resources in R&D activities (Probit column), firm size has a positive and significant impact on this probability. This increase with size is a well-documented fact in the empirical literature. After controlling for size and sector, it appears that the probability of doing R&D also increases with the degree of market share. As suggested by Crepon et al (1998), this is a new link that they also found in their analysis of French firms.

The analysis with R&D intensity equation shows that firm size has no significant impact on this variable, suggesting that the elasticity of R&D expenditure to size is one. This is consistent with previous work in this line of research.⁷ However, contrary to the French case, in our study with Chilean firms neither market share nor firm diversification has a significant impact on R&D intensity. It seems that these two variables only affect the decision to enroll in R&D activities but not on the size of it. This last fact is also consistent with previous work summarised in Cohen and Levin (1989) and Cohen (1995).

Regarding the innovation equation, current R&D expenditures has no significant impact on current sales weighted innovations, once controlled by size and sector. This result can be explained by the dynamic nature of the innovation process where is difficult to expect instant success related to R&D activities. However, in the case of Chile, it appears that bigger firms tend to have more success in their innovative sales compared to their smaller counterparts.

[Table 2 about here]

Finally, estimates for the productivity equation shows that there exist constant returns to scale and a physical capital elasticity of about 0.7. However, current sales weighted innovations which proxies innovation efforts has no statistical significance in the explanation of productivity of

⁷ Known as Stylised Facts 2 and 3 under Cohen and Kepler's (1996) terminology.

firms.⁸ This result may be due to the flow characteristic of the Chilean innovation data.

4.2 Results of the Extended Model

The ALS estimates of the extended version of the model are presented in Table 3. We have included six additional rows in the research and innovation equations (for the weak, moderate and strong type of answers) which define our demand pull and technology push indicators, and two additional rows in the productivity equation for our two skill composition variables : the share of engineers and administrators relative to other categories of employees.

With respect to the variables already discussed, they do not change dramatically between either specification. The only exception is firm size, which appears to be significant in the productivity equation but not in the innovation one.

Both market demand and technological opportunities, measured here by demand pull and technology push indicators, appear to have positive effects on the firm's probability of engaging in R&D activities and their intensity. However, only the proxies for technological opportunities have a significant effect in the Chilean plants. In the case of research intensity, these effects tend to increase across the type of answers as in the French case. Surprisingly, in the case of the probability equation, technology push variables are inversely related with probability. That is, the likelihood that firms allocate resources to R&D activities does not increase if managers perceive that technological opportunities are important, moreover it may decrease. These estimators are, however, positive and statistically significant. In the case of the innovation equation, neither demand-pull nor technology push indicators have a significant effect on the sales weighted innovations.

Finally, our two skill composition variables are significant and with large positive coefficients in the productivity equation.⁹ Especially

⁸ As can be observed in Table II.6 in the Appendix II, a similar result is obtained when R&D expenditure is used instead.

⁹ Also they are not correlated with knowledge capital proxies when comparing both specifications, suggesting no complementarity nor substitution between these factors.

administrative workers, and in a second place engineers, make a massive difference in their impact on the plant's productivity compared with other type of workers. This is obviously reflected in their relative salaries.¹⁰

[Table 3 about here]

5 Conclusions

One of the main characteristics of empirical studies on R&D and innovation is the particular structure of the empirical models adopted. Most of the research in this area has been conducted by using a simple equation relating a measure of research effort to firm characteristics and/or market structure.

There exists several biases that arise by using single-equation models. Probably the most important is the selectivity one which is the result of using samples including firms with no reported R&D expenditure. Recent work has tackled this problem by estimating a system of equations where one equation deals with the selectivity issue and the other with the research intensity. A Generalised Tobit procedure is the adequate tool to estimate these systems in a consistent and efficient manner. However, we are only dealing with part of the whole innovative process that takes place at the level of the firm. One of the main objectives of this paper is to explore the links between R&D and innovation, and the impact of them on firms' productivity for a LDC case.

Results show that when all the steps in the innovative process are incorporated - R&D, introduction of new products and processes, and firm productivity - most of the traditional hypotheses of firm characteristics and market structure are validated under this broader set-up.

Indeed, in the case of Chile, firm size is related to the probability that firms are engaged in research activities. However, size is not related to the amount of resources allocated for these activities once controlled for sectorial differences, suggesting a constant return to scale in research investments. Results also show that technological opportunities do play a

¹⁰ See INE (1998).

major role in research activities specially when innovative ideas are embedded in new machinery and output.

There is a constant return to scale in the productivity equation and after including labour skills in the estimation of the productivity, both engineers and administrative shares have a positive and significant effect.

The econometric methods used take into account not only the particular characteristics of the data such as truncation, interval and censored but also the selectivity and simultaneity problems that are embodied in this kind of empirical exercise.

However, some of the findings of our study of the Chilean case were unexpected. Among them the fact that neither research expenditure nor innovation has a significant impact on innovation sales and productivity respectively. These results can be explained partly by the implicit assumption in the model that there are no lags between the implementation of innovations and impact on productivity, a subject widely discussed in the literature. However, a more plausible explanation is related with the fact that productivity is measured as value added per worker. If innovation is mainly related with embodied technical change, then this effect is not captured in this model. Results in related work¹¹ suggest that machinery is related with the probability to report R&D expenditures but we do not have the data to test this hypothesis for productivity.

We intent to tackle this problem in the near future when more information for the Chilean firms will be available.

6 References

Benavente, J.M. (2002): "Determinants of Industrial Research and Innovation : The case of Chile ", unpublished DPhil Thesis, University of Oxford.

Benavente, J.M. and G. Crespi (1999): "Informe Final. Encuesta sobre la Innovacion Tecnologica en la Industria Manufacturera Chilena", manuscript, University of Chile.

Bera, A., Jarque, C. and L.F. Lee (1984): "Testing the normality

¹¹ Benavente (2002)

assumption in limited dependent variable models", *International Economic Review*, 25, 563-578.

Cohen, W. (1995): "Empirical Studies of Innovative Activity" in P. Stoneman (ed.) *Handbook of the Economics of Innovation and Technological Change* (Blackwell Handbooks in Economics).

Cohen, W. and R. Levin (1989): "Empirical Studies of Innovation and Market Structure" in R. Schmalensee and R. Willig (eds.) *Handbook of Industrial Organization* (North Holland).

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

Crepon, B., Duguet, E. and J. Mairesse (1998): "Research, Innovation, and Productivity : An Econometric Analysis at the Firm Level". NBER Working Paper 6696.

Gourieroux, C. and A. Monfort (1996): *Statistics and Econometric Models* (Cambridge University Press).

Francois, J.P. (1995): "Innovation technologique dans l'industrie. Problematique et quelques validations empiriques", *Estadística y Economía*, 10, Santiago de Chile.

INE (1998) : "Ingreso de Hogares y Personas. Encuesta Suplementaria de Ingresos " , *Instituto Nacional de Estadísticas*, Santiago de Chile.

Lee, L.F. (1981): "Simultaneous equations models with discrete and censored variables " , in C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press).

Pagan, A and F. Vella (1989): "Diagnostic test for model based on individual data : a survey", *Journal of Applied Econometrics*, 4, S29-S59.

Rosenberg, N. (1974): "Science, innovation and economic growth " , *Economic Journal*, 84, 90-108.

Schmookler J. (1966) *Invention and Economic Growth* (Harvard University Press).

Schumpeter J.A. (1934) *The theory of economic development* (Harvard University Press).

Schumpeter J.A. (1942) *Capitalism, Socialism, and Democracy* (Harper and Row).

Appendix I

I.1 The Survey

The data used in the first part of this study comes from the first Technological Innovation in the Chilean Manufacture Survey held during 1998. It was a joint work between the National Institute of Statistics (INE) and the Ministry of Economics. The survey was conducted according to the OECD instructions defined in the Oslo and Frascati Manuals.

In general, the survey was designed to capture qualitative variations in the innovation activities of firms. The main objective was to identify the factors underlying these changes and to evaluate the efficiency of public policy. Therefore, the survey contains only a limited number of measures about the levels of innovation activity within firms with the only exception of expenditures on Research and Development and other innovation activities like patents, royalties and training in R&D. However, because the sample of firms was obtained from the universe of firms contained in the Annual Industrial Survey (ENIA), other quantitative characteristics of the firms can be included and analysed (e.g. employment, exports, investment).

A second feature of the design of the survey is the procedure to estimate variations in the innovation activity. A set of questions was asked to capture perceptions from engineers and production managers of the plants. Therefore, the *measurement* of the changes in the innovation dynamics was made by the responsible of the production units. To standardise the answers they were ordered according to the intensity or importance given to each question in those cases where an innovation was present. The rank varies from 1 to 4 (from lower to higher importance).

A third aspect to highlight is that although the survey contains only a single observation in time, some dynamic behaviour can be obtained from it. The questions were designed to capture information on productive performance during the last three years rather than the year in which the survey was conducted.

There are some limitations to be taken into account. First, there is room for arbitrariness in the intensity scaling and its interpretation. Second,

the problem of aggregating answers that are qualitative in nature. In fact, from an answer of 4 (which represents that the phenomena was very intense or very important in the firm) cannot be automatically inferred that the phenomena have taken the same characteristics in all the firms who gave 4 as an answer for the same question. An answer of 4 in two firms to the same question have to be interpreted that the phenomena has been perceived as important or intensive in each firm despite differences between each case.

Finally, the evaluation of variations rather than levels of innovative activity make very difficult the international comparisons.

1.2 The sample of plants

The information referred to plants rather than firms since the survey aimed to detect innovative phenomena developed at a plant level. Nevertheless, in the cases of firms with more than one plant, the survey was applied to all the plants of the same firm in order to capture the potential externalities and links between the plant of the same firm. A probabilistic and stratified sampling procedure, considering the size by value added, was performed for the study.

The design considered two groups of plants. The first is named *forced inclusion strata*. The inclusion criteria here is plants with more than 2% of the value added or exports or investment of his productive sector (2 digit ISIC). The second group is randomly selected. The selection was made in each size strata with a probability proportional to the value added of each plant. The final sample included 488 plants where the method chosen to select the sample guarantees a good representation of the universe with a sampling error of 5.1% at 95% of confidence.

1.3 The data

Traditional variables

The information on the firm current accounts and balance sheets, and on the number of employees comes primarily from the ENIA surveys (Encuesta Nacional Industrial Anual). From them we have constructed the firm value added, its fixed assets gross bookvalue and its total number of employees

(average over the year), and we have computed labour, productivity and physical capital intensity as l_i = number of employees, q_i = value added per employee and c_i = physical capital per employee. These variables are expressed in logarithm in our estimations.

The information about the distribution of employees comes from the same source. The data allows to differentiate between administrative and engineers at a managerial level from blue-collar workers at the floor level and the rest of the employees.

Average firm market share and diversification indexes are computed from the same ENIA survey. This survey gives detailed information about the decomposition of firm's sales in all its different lines of business up to a Chilean equivalent of 8 digit ISIC. In our estimations we have defined $S_{i,k}$ as the sales of firm i for its product k in the industry segment or market k .

$$S_i = \sum_k S_{i,k} \text{ and } S_k = \sum_i S_{i,k}$$

are respectively the overall sales of firm i (overall its products) and overall sales on market k (overall firms) without including its exports. The market share $s_{i,k}$ of firm i on market k and the share of product k on firm i total sales are thus equal to:

$$s_{i,k} = \frac{S_{i,k}}{S_k} \text{ and } b_{i,k} = \frac{S_{i,k}}{S_i}$$

Then for each diversified firm i we can define the weighted average market share s_i^w and the diversification index d_i as :

$$s_i^w = \sum_k b_{i,k} \times s_{i,k} \text{ and } \frac{1}{d_i} = h_i = \sum_k b_{i,k}^2$$

with d_i being the inverse of the Herfindahl concentration index h_i of the firm sales. For example, a non-diversified firm, i.e. with only one k , we have $s_i^w = s_i$ and $d_i = h_i = 1$. The diversification index d_i for the firm i can be interpreted as the equivalent number of product lines with equal sales. The higher this index the higher the diversification of the firm. In the model we introduce variables s_i^w and d_i in logarithms.

Finally we include in all equations of the model a full set of 8 industry dummies : $S_i^1, S_i^2, \dots, S_i^8$. These are defined at a higher level of

classification (2 digits ISIC) than the average market share and diversification variables, on the basis of the firm main industrial activity.

Research and Innovation data

Information about R&D expenditures is taken from the “Encuesta de Innovación Tecnológica en la Industria Manufacturera Chilena”. Plants were requested to inform about their R&D expenditures during 1998 and 1997 translated in 1998 Chilean pesos using the Frascati methods. This survey has two versions, the first in 1995 and the second in 1998. Given the characteristics of the sample design, only few firms (less than 200) were interviewed in both occasions so larger time series of research and innovative activities were difficult to obtain. In our model, g_i is a research dummy equal to 1 if firm i has reported R&D expenditures during 1997 and 1998 and 0 otherwise. On the other hand k_i is the log of the R&D expenditures of 1998.

The share of innovative sales is also obtained from the same survey. This variable is based on a question asking innovating firms what percentage of their 1998 sales is imputable to new products and/or new processes launched between 1996 and 1998. Firms answered on an interval scale : 0-10%, 10-30%, 30-70% and more than 70%.

Firm's demand pull and technology push indicators are respectively based on the average answer to the following questions: do you think that the ideas for innovation come from either research activities performed together with clients and suppliers, other firms or simply copy?, and do you think that innovations comes from concepts and ideas embodied in your purchases of new machinery, the use of new inputs and/or the consultation of books, periodicals or visiting expositions ?. Questions are ranked in a scale between one and four, giving a qualification of weak for those answers equal to 1, moderate for 2 and 3, and strong for only those answered with a value of 4.

Appendix II

An assessment of the biases likely to arise in innovation and productivity studies

One of the main objectives of this study is to correct for selectivity and simultaneity biases commonly found in applied innovation studies. We have also taken into account the main characteristics of the available data like censoring and truncation. We also intend to assess the magnitude of these biases while using more standard econometric methods. In this section we present results of the estimation of the same group of equations using ordinary least squares (OLS), two stage least squares (2SLS), maximum likelihood (ML) and first step and second step asymptotic least squares (ALS) estimators. We only perform this exercise for the basic specification of our model to the sample of innovative plants and when appropriate for the sub sample of R+D performing firms.

Research Equations

Table 4 compares the OLS estimates for the research intensity equation for the subsample restricted to the R&D performing firms to the corresponding ML estimates for the generalised Tobit specification. As can be seen they differ substantially. Moreover, if OLS estimation was used instead we would have concluded that market share is an important explanation in the R&D intensity of firms and that the elasticity of R&D to firm size is bigger than one, which both are not the case.

Innovation Sales Equation

Before discussing these results, is important to notice that for OLS and 2SLS estimation we have assumed that the value of innovative sales was simple the interval centre, that is 5%, 20%, 50% and 85% respectively.

As can be observed from Table 5 there are problems with selectivity given the differences in the value for the estimates between the two samples. However, it has to be noted that particularly in the case of 2SLS estimates are very imprecise. On the other hand, simultaneity problems can exist

while comparing results from ML estimation with ALS. For example, note that in the case of firm size, this coefficient is not significantly different from zero under ML but is so under ALS. It seems that ML is not taking into account the effect that firm size does through the probability to perform R+D activities into innovation sales.

Productivity Equation

The first interesting observation of the results presented in Tables 6 and 7 is that capital is always significant and with an elasticity about 0.7 in all the equations. The exception is again the 2SLS, which, as in the French case, estimates using this technique are quite imprecise. Table 6 also shows that there exist selectivity problems when only the sub-sample of R+D performers firms is considered. Estimators differ substantially between both samples.

In none of the estimations, R+D expenditures and/or innovation sales has a significant effect on firm s productivity. This confirms the results obtained with ALS, which corrects for potential problems due to selectivity and simultaneity, where knowledge-related variables are not important in explaining productivity changes, at least contemporarily, in the Chilean manufacturing industry.

Table 1. Descriptive statistics of the Innovation Survey and its expansion to the whole of Chilean Industry. Selected Indicators.

Sample	Sample Data (488 plants)		Sample adjusted by weights	
	Mean	%	Mean	%
R&D expenditure per employee* F_i (1998)	292	44.7	280	15.2
Number of employees L_i (1997)	291	100	86	100
Market share (%) $S_i \times 100$ (1997)	0.34	100	0.13	100
Equivalent number of activities D_i (1997)	2.15	59.2	2.17	98.5
Introduction of innovation (%) Y_i (1998)	29.5	id.	17.9	id.
Value added per employee* Q_i (1997)	42,929	100	10,497	100
Physical capital per employee* C_i (1997)	61,198	100	12,972	100
Engineers/Employment (%) $E_i \times 100$ (1997)	20.72	100	18.34	100
Administrative/Employment (%) $A_i \times 100$ (1997)	3.39	79.1	3.14	32.4

* Thousand of 1998 Chilean Pesos CHP

Table 2. Results for Basic Model

Left hand Variables :

Logarithm of research expenditures per employee (k_i)

Logarithm of innovation intensity (t_i)

Logarithm of value added per employee (q_i)

(asymptotic and tests standard errors between parentheses)

Variable	R&D		Innovation	
	Probit	Tobit	Intensity	Productivity
R&D expenditure per employee k_i	-	-	-0.076 (0.199)	-
Innovation intensity t_i	-	-	-	0.170 (2.072)
Number of employees l_i	0.182 (0.055)	0.082 (0.129)	0.142 (0.029)	0.100 (0.185)
Market share s_i	0.139 (0.049)	0.195 (0.121)	-	-
Equivalent number of activities d_i	-0.025 (0.130)	-0.078 (0.276)	-	-
Physical capital per employee c_i	-	-	-	0.686 (0.044)
Statistical Test (Pagan-Vella,Bera-Jarque-Vella)	1.57 (0.45)	0.103 (0.92)		
Optimal asymptotic least squares with 8 industry dummies Sample of 438 plants.				

Table 3. Results for the Extended Model

Left hand variables :
 Logarithm of research expenditures per employee (k_i)
 Logarithm of innovation intensity (t_i)
 Logarithm of value added per employee (q_i)
 (standard errors between parentheses)

Variable			R&D		Innovation	
Statistics			Probit	Tobit	Intensity	Productivity
R&D expenditure per employee		k_i	-	-	-0.181 (0.344)	-
Innovation intensity		t_i	-	-	-	0.179 (0.113)
Number of employees		l_i	0.077 (0.059)	-0.115 (0.130)	-0.011 (0.064)	0.140 (0.041)
Market share		s_i	0.126 (0.052)	0.143 (0.116)	-	-
Equivalent number of activities		d_i	-0.097 (0.132)	-0.153 (0.262)	-	-
Physical capital per employee		c_i	-	-	-	0.670 (0.021)
						0.652 (0.177)

Engineers/Personnel E_i	-	-	-	5.952 (1.459)
Administrative/Personnel A_i				
Demand Pull :	0.191 (0.273)	0.035 (0.542)	0.360 (0.280)	-
Weak (δ_i^1)	0.286 (0.234)	0.157 (0.451)	0.085 (0.247)	-
Moderate (δ_i^2)	0.292 (0.220)	0.111 (0.430)	0.335 (0.227)	-
Strong (δ_i^3)				
Technology Push :	0.980 (0.312)	1.778 (0.641)	0.890 (0.671)	-
Weak (τ_i^1)	0.904 (0.255)	2.011 (0.540)	1.139 (0.735)	-
Moderate (τ_i^2)	0.895 (0.244)	2.223 (0.522)	1.402 (0.795)	-
Strong (τ_i^3)				

Optimal asymptotic least squares with 8 industry dummies
Sample of 438 plants.

Table 4. Research and Development Equation

Dependent variable :

Logarithm of research and development expenditures per employee (k_i).
(standard errors between parentheses)

Variable	l_i	s_i	d_i
Generalised Tobit (maximum likelihood)			
Probit part	0.182 (0.055)	0.139 (0.049)	-0.025 (0.130)
Tobit part	0.082 (0.129)	0.195 (0.121)	-0.078 (0.276)
OLS	0.297 (0.105)	0.183 (0.079)	-0.030 (0.234)

All regressions include 8 industry dummies.

Correlation among the residuals = 0.830 (0.057).

Table 5. Innovation Intensity Equation

Dependent variable :

Logarithm of innovation percentage in sales (t_i).

(standard errors between parentheses)

Variable	k_i	l_i	g_i
All Observations (438 obs.)			
Ordinary least squares *	-0.030 (0.046)	0.177 (0.050)	0.360 (0.223)
Two stage least squares	-2.766 (5.142)	-0.077 (0.460)	12.830 (23.432)
Maximum likelihood +	-0.071 (0.051)	0.073 (0.056)	0.594 (0.247)
Asymptotic least squares (first step : OLS)	0.105 (4.750)	0.060 (1.073)	*
Asymptotic least squares (second step : GLS)	0.133 (0.241)	0.060 (0.061)	*
Positive R+D (197 obs.)			
Ordinary least squares *	0.036 (0.056)	0.190 (0.086)	*
Two stage least squares	2.539 (5.046)	0.134 (0.432)	*
Maximum likelihood +	-0.093 (0.057)	-0.007 (0.089)	*

All regressions include 8 industry dummies.

Instruments for 2SLS : l_i, s_i, d_i and the 8 industry dummies.

* The logarithm of the interval centre was taken as the dependent variable.

+ Ordered probit with known thresholds.

Table 6. Productivity Regression with R&D expenditures

Dependent variable :

Logarithm of value added per employee (q_i).

(standard errors between parentheses)

Variable	k_i	c_i	l_i	g_i
All Observations (438 obs.)				
Ordinary least squares	0.058 (0.037)	0.693 (0.050)	0.129 (0.048)	-0.159 (0.184)
Two stage least squares	-2.603 (4.836)	1.057 (0.670)	-0.325 (0.812)	11.943 (21.986)
Positive R+D (197 obs.)				
Ordinary least squares	0.131 (0.040)	0.686 (0.068)	0.197 (0.078)	*
Two stage least squares	8.195 (318.88)	-4.057 (187.58)	2.372 (85.932)	*

All regressions include 8 industry dummies.

Instruments for 2SLS : l_i, c_i, s_i, d_i and the 8 industry dummies.

Table 7. Productivity Regression with Innovation Intensity
 Dependent variable :
 Logarithm of value added per employee (q_i).
 (standard errors between parentheses)

Variable	t_i	c_i	l_i
All Observations (438 obs.)			
Ordinary least squares *	0.060 (0.048)	0.696 (0.048)	0.121 (0.043)
Two stage least squares *	0.833 (0.821)	0.619 (0.083)	0.009 (0.126)
Asymptotic least squares (first step : OLS)	0.362 (35.93	0.705 (1.000)	0.104 (2.663)
Asymptotic least squares (second step : GLS)	2) 0.032 (2.099)	0.689 (0.044)	0.109 (0.188)

All regressions include 8 industry dummies.

Instruments for 2SLS : l_i, c_i, s_i, d_i and the 8 industry dummies.

* The logarithm of the interval centre was taken as the right-hand variable.