Will a BERD model fly?
Estimating aggregate R&D expenditure using a micro model*

Richard Fabling

June 2008

DRAFT paper
Prepared for ESAM'08, Wellington

* The author wishes to thank Julia Gretton for extremely able research assistance in preparing the data, and Dave Maré, Lynda Sanderson, Hamish Hill and Gary Dunnet for helpful comments on the paper. Any remaining errors are the responsibility of the author. The views in the paper are those of the author and should not be attributed to the Reserve Bank of New Zealand.

Will a BERD model fly?
Estimating aggregate R&D expenditure using a micro model

Richard Fabling[1]

Abstract

This paper tests the feasibility of using a microeconometric model to construct a consistent measure of Business Expenditure on Research and Development (BERD) over time. This approach is motivated by an attempt to compensate for changes to survey-design that make it difficult to interpret recent growth in R&D – a key economic indicator of the innovativeness of the economy. We begin by estimating a two-stage selection model of the determinants of R&D investment decisions for private-for-profit firms in New Zealand. The first stage yields estimates of whether a firm performs R&D, while the second stage estimates R&D intensity taking into account the determinants of the decision to make the investment. Using Statistics New Zealand's prototype Longitudinal Business Database, we are able to consider a wide range of potential determinants of R&D activity including ownership, industry, balance sheet structure and prior performance. After appraising the appropriateness of the R&D model, we use it to predict expected R&D expenditure for the population of firms and, from this, derive a time series for aggregate BERD.

JEL classifications: C82; D21; O30

Keywords: research and development; firm behaviour

[1] Reserve Bank of New Zealand (Richard.Fabling@rbnz.govt.nz); Motu Economic and Public Policy Research

DISCLAIMER

This research uses data that was accessed while the author was on secondment to Statistics New Zealand in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Act are allowed to see data about a particular business or organisation. The results of this work have been confidentialised to protect individual businesses from identification. The analysis and interpretation of these results were undertaken while the author was at the Reserve Bank of New Zealand. The opinions, findings, recommendations and conclusions expressed in this report are those of the author. Statistics NZ, the Reserve Bank of New Zealand, and Motu take no responsibility for any omissions or errors in the information contained here.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information is published or disclosed in any other form, or provided back to Inland Revenue for administrative or regulatory purposes. Any person who had access to the unit-record data has certified that they have been shown, have read and have understood section 81 of the Tax Administration Act 1994, which relates to privacy and confidentiality. Any discussion of data limitations or weaknesses is not related to the data's ability to support Inland Revenue's core operational requirements.

Statistics NZ protocols were applied to the data sourced from the New Zealand Customs Service and the Foundation for Research, Science and Technology. Any discussion of data limitations is not related to the data's ability to support these government agencies' core operational requirements.

1 Motivation

Research and development (R&D) is generally believed to be a major mechanism by which firms can improve or create products and production processes. R&D can create new market opportunities and temporary monopoly power, raise profits and improve productivity for the firms concerned. R&D is further considered to have wider benefits than those that are captured by the instigating firm. Since it may be difficult for firms to fully protect the knowledge they create, a positive spillover may occur to other parts of the economy. Knowledge generated through R&D (and other firm investments) can thus be seen as a possible cause of long-term economic growth, where future knowledge creators build on the work of those before them, increasing the knowledge "stock" on which economic production is founded (Romer 1990). For this reason, measured R&D in the whole economy, and particularly the business sector, is seen as a key driver of economic performance (eg, OECD 2006, 2007; Furman, Porter and Stern 2002).

The presence of economic benefits that cannot be completely captured by the individual firms investing in R&D creates a strong case for governments to compensate R&D performers. International estimates of the private and social returns to R&D have stimulated many researchers and policymakers to attempt to describe or explain the R&D performance of the New Zealand economy (eg, Mazoyer 1999; Davis 2006; Hall and Scobie 2006; Smith 2006; MoRST 2006a; Crawford *et al.* 2007; Johnson, Razzak and Stillman 2007). Partly as a result of this effort, the New Zealand government recently announced the decision to adopt a tax credit, directly compensating R&D-performing firms for the broader returns from their investment decision.

From an official statistics perspective, the story is mixed on New Zealand's Business Expenditure on R&D (BERD) performance. Figure 1 shows Organisation for Economic Co-operation and Development (OECD) comparative figures for the evolution of R&D intensity over the last decade. According to these statistics, New Zealand's <u>growth</u> in R&D expenditure has been quite strong (relative to Gross Domestic Product), with an apparent doubling in the ratio over the decade. However the <u>level</u> of the ratio of R&D to GDP has been, and continues to be, weak in a cross-country sense.[2] The theoretical arguments discussed briefly above suggest we should care greatly about our historic R&D performance because it is the accumulated (depreciated) stock of knowledge generated by R&D that matters for production. That is, in the same way that the purchase of a computer yields flows of capital services over several years, outputs from investment in R&D can be seen as contributing to the ongoing production of goods and services until a subsequent "invention" renders the current knowledge obsolete.

Potentially, New Zealand's historical BERD is not as weak as reported. Specifically, the methodology underlying the collection of R&D Survey data has changed substantially over the last ten years, evolving from a census of known R&D performers, to a mixed list-sample survey approach.[3] While this series of changes represent important methodological

---

[2] Though Crawford *et al.* (2007) provide evidence that our poor relative performance can largely be explained by slowly-evolving (or permanently fixed) features of the NZ economy, particularly the average firm size, industry composition and distance from major R&D-performing countries.

[3] The R&D Survey is a joint survey with the Ministry of Research, Science and Technology (MoRST) and is carried out biennially by Statistics New Zealand. It measures the level of research and development activity, employment and expenditure by business sector enterprises, government departments, government-owned trading entities, and universities. R&D data is also collected through the Business

improvements in the estimation of BERD, every step towards a full sample survey has uncovered numerous firms that were not previously considered R&D performers. Such discoveries have meant that it is difficult to compare estimated measures of BERD over time. This is recognised by Statistics New Zealand, who advise "caution… when making comparisons between 2004 results and previous survey periods" (Statistics New Zealand 2004) and that "R&D expenditure figures in the 2002 reference year are not directly comparable with those collected in 2000 and previous years because of changes in the methods employed to identify enterprises undertaking R&D activity" (Statistics New Zealand 2002).

The problem of inter-temporal consistency is addressed by the Ministry of Research, Science & Technology in their *Decade in Review* publication (MoRST 2006a). MoRST's approach to creating a consistent time series was to convert aggregate values with sample-survey components into "list-based measures" and to harmonise the population coverage to exclude firms with less than ten employees (MoRST 2006b). This reduced the observed measure of BERD in 2002 and 2004 (the final year the study covered) by roughly $100 million per year, or approximately 15 percent of the total. These adjustments also materially affect the estimated number of firms performing R&D, with that count dropping by over 1,000 firms (that is, by more than half). Overall, this harmonisation effort challenges the perception of recent rapid growth in New Zealand R&D, attributing a very large share of that apparent growth to changes in survey population and design. Figure 2, sourced from MoRST (2006b), provides a graphical summary of this point. Their methodology suggests a downward revision of nominal BERD growth from roughly 19 percent to around 10 percent per annum.

This paper takes an alternative approach, starting from the premise that the recent sample survey approach adopted by Statistics New Zealand provides strong evidence that aggregate BERD is seriously underestimated historically. Like the MoRST approach, this assumption also implies that aggregate BERD has been growing at a much slower rate than previously reported. However, in contrast to that earlier work, our assumption also implies that the accumulated investment in R&D – the stock of R&D knowledge generated in New Zealand – is much higher than previously thought.[4]

In order to construct a consistent measure of BERD for all employing private-for-profit firms, this paper uses an econometric two-stage selection model to estimate the relationship between R&D activity and a wide range of firm characteristics including prior R&D behaviour, firm performance and market structure. The model is then used to predict R&D for every firm in the sector, yielding an estimate of BERD. Section 2 outlines the motivation of the model based on the international literature. Section 3 describes the data underlying the estimation, focusing on the breadth of information available from Statistics New Zealand's prototype Longitudinal Business Database (LBD) as well as the multiple sources of R&D data available at the micro level. Section 4 presents model estimates and implied aggregate BERD values, while Section 5 draws out conclusions from these findings, particularly in regard to the existing empirical literature that makes use of BERD official statistics.

Operations Survey (BOS), though not for the purpose of producing official BERD statistics. The relationship between BOS and R&D Survey data is discussed later in the paper.

4 This does not necessarily imply that New Zealand's relative historical performance in either levels or growth rates is understated. A detailed harmonisation of data methods across countries (taking account of sample design changes elsewhere) would be the only fair basis on which to make that comparison. We are unaware of any study that has attempted to harmonise international R&D data to that extent.

2 Model

The international literature suggests several reasonably strong hypotheses related to the decision to perform R&D. Smaller firms are found to be less likely to conduct R&D but, once this decision is made, firms in the same industry invest at similar intensities.[5] This stylised fact suggests an econometric two-stage selection model (ie, Heckman) is appropriate to consider what influences R&D investment decisions. Such a model allows potential determining factors to have different influences on the decision to do R&D and the scale of the investment decision. The first stage estimates whether a firm performs R&D, while the second stage estimates R&D intensity conditional on R&D being performed. The key to estimating such models is to find variables that might reasonably influence the decision to perform R&D, but not affect how much R&D is done. In the absence of variables in the selection equation that do not appear in the intensity equation, the two-stage model is identified purely from functional form assumptions.

We now present a quick summary of potential R&D determinants based on two prior literature reviews by Symeonidis (1996) and Becker and Pain (2008). With varying degrees of success various authors have considered – among other things – the following loosely categorised determinants of R&D investment decisions:

Firm size: The importance of size to the investment decision is often seen as reflecting a combination of economies of scale or scope in R&D, portfolio effects that mitigate risk (eg, being able to afford a diversified portfolio of R&D projects), superior access to finance, or a greater ability to generate returns from innovations (eg, because existing client bases are larger).[6]

Internal financing: If capital markets are incomplete, then the ability of firms to finance "lumpy" and potentially risky R&D might be compromised. Past profits are often used as an indicator of the ability to finance projects internally.[7]

Competition/market power: Recent empirical effort has focused on whether competition has a non-linear effect on innovation effort. For example, oligopolistic markets may provide the optimal conditions for investment because current market players are big enough to make large-scale R&D investments, and yet they are still sufficiently incentivised to compete for the market (Aghion *et al.* 2005).

Location: Proximity to other sources of R&D output/input may be beneficial based on, among other arguments, the potential for spillovers from the effort of other firms to be transmitted by personal interactions or labour mobility. The important role that distance has in inhibiting the diffusion of potential knowledge transfer has been considered in many contexts including citations (eg, Jaffe, Trajtenberg and Henderson 1993) as well as the determinants of aggregate R&D expenditure (eg, Crawford *et al.* 2007).

Technology characteristics: It may be that some industries have different product cycles and the demand to produce new technologies differs across these markets. While some studies try to directly control for technological characteristics, most studies rely on industry dummies to control for this effect. To give a sense of the importance of industry differences in R&D effort, Mazoyer (1999) decomposed the difference between New Zealand BERD over GDP

---

5 R&D intensity is normally measured as the ratio of R&D expenditure to sales, value-added or total employment.

6 This last ignores the potential dynamic effects of successful R&D efforts on firm size.

7 Though Symeonidis (1996) notes that past profits might also represent expectations of future profits or the success of past R&D efforts.

and the OECD average into the share attributable to New Zealand's unique industry mix and a share related to differences from international industry norms for R&D. She found that roughly half the difference in New Zealand R&D intensity can be explained by industry composition. Along similar lines Crawford *et al.* (2007) found that the share of employment in primary industries was important in explaining cross-country differences in R&D.

Government support: As discussed earlier, there are good theoretical arguments why governments should support private R&D. In the empirical literature, the potential impact of government policy has been considered from a positive perspective, through say tax credits and public university collaboration, and also as potentially crowding out private effort (David, Hall and Toole 2000 reviews this literature).

Human capital: R&D is labour- and skills-intensive so that the supply of adequately talented people is a likely prerequisite of being able to perform R&D. In New Zealand, salaries and wages constitute roughly half of BERD (Statistics New Zealand 2006).

International engagement: Arguments can be made either for or against certain types of international connections being positive for R&D activity. Exporting firms may be more likely to do R&D because they are exposed to other markets and, therefore, a larger knowledge base, or they may simply face greater competition. Foreign-owned firms, on the other hand, may do more or less R&D in New Zealand depending on where the innovation functions of the multi-national firm are best located. That location decision could be driven by, say, synergies with other business functions (such as production facilities or head offices) or the relative costs and benefits of doing R&D in the various jurisdictions in which the firm operates (eg, because of differing tax incentives).

Persistence: R&D may involve both set-up and close-down costs so that firms may be inclined to continue investing once they start. Examples include hiring and firing costs, capital investments associated with equipping laboratories, or purchases of existing intellectual property rights.

Macroeconomic conditions: Key variables that have been considered in the past include interest rates – which might influence the investment hurdle rate for undertaking projects – and exchange rate movements, which potentially capture exogenous variation in import competition.

Studies often only include manufacturing firms in their sample. We have broader coverage (discussed in the next section) and, because of the technology characteristics arguments, it may be important to consider the interaction between industry and other variables. We should be particularly concerned whether some of our right-hand side variables may differentially influence decisions in "high" and "low" technology industries. We are keen to distinguish between markets where R&D is a "rule of entry" and the norm (eg, advanced manufacturing), and others where R&D is a rare undertaking (eg, retail trade).

Because our model is to provide estimates of aggregate BERD, we are restricted to considering right-hand side variables that are held by our population of firms. Fortunately, our dataset is extremely rich, even when restricted to data that most firms in the population have. We now turn to a discussion of that data.

3 Data

The LBD contains a broad range of firm performance data from both survey and administrative sources. The core administrative data on the LBD consists of the Longitudinal Business Frame (LBF)[8] with goods and services tax (GST) returns,[9] financial accounts (IR10),[10] and company income tax returns (IR4)[11] provided by Inland Revenue; information on employers and employees aggregated to the firm level, sourced from the Linked Employer-Employee Dataset (LEED);[12] and shipment-level merchandise export and import data provided by the New Zealand Customs Service.[13] The coverage of the database is extremely broad and, at a minimum, covers all firms in the economy with at least $40,000 in taxable income. For this paper we also make use of three sample surveys that are included in the LBD: the Annual Enterprise Survey;[14] the R&D Survey; and the Business Operations Survey (BOS). Fabling *et al.* (2008) and Statistics New Zealand (2008) go into greater detail on the full contents of the database, and more information on the individual data sources is available from Statistics New Zealand's website.

In this paper we focus on the private-for-profit sector of the economy. This population is defined by excluding government, private households and non-profit business types (such as trusts). We further restrict our population to exclude non-employing firms because we expect some labour input to be necessary to conduct or commission R&D. Having made that restriction, we specify our R&D intensity on a per employee basis. We have all the necessary data available to compute an estimate for BERD for a six-year period from 2001-2006.[15] Table 1 provides summary statistics on the composition of our population. Perhaps the most interesting feature of Table 1 is the fact that our count of firms is very stable over the six years. At face value, this result appears inconsistent with previously published figures, which tend to show the population of private-for-profit firms growing steadily over time (Fabling *et al.* 2008). The apparent difference in results is reconciled by the fact that our population only includes employing firms – most of the growth in firm numbers has come through the expansion of the population of non-employing firms.

Because of the unique potential of the LBD, we are able to consider a wide range of potential determinants of R&D activity including ownership type, industry, balance sheet structure and prior performance. In particular, we can directly control for many of the factors that firm size may proxy for in other studies. Before turning to those measures, we first discuss our R&D

8 The LBF is a variant of Statistics New Zealand's sampling frame and contains longitudinal information (eg, industry, ownership type, and sector) on firms.

9 GST data include information on sales and purchases. We use this data after it has been processed to create Statistics New Zealand's Business Activity Indicator (BAI). The primary benefit we gain from using BAI data is that group-filed returns have been apportioned across GST group members.

10 IR10 data is essentially a set of company accounts composed of a statement of financial performance and financial position.

11 IR4 returns are declarations of taxable income for companies and, as such, include variables on overseas income, interest and dividends, and income from "business or rental activities".

12 LEED data is constructed by Statistics New Zealand from Inland Revenue tax data, notably Pay-As-You-Earn (PAYE) returns for employees.

13 Customs data is linked to the LBF initially via probabilistic matching on names and addresses, with subsequent manual matching for remaining unmatched large-value Customs clients.

14 AES is Statistics New Zealand's primary data source for the production of National Accounts and, as such, is the benchmark dataset for estimation of value-added.

15 While we have full coverage data in 2000, our model makes use of lagged financial information and, at present, there is no imputation in the database for 1999.

measure since we have three potential sources for this. While the R&D Survey is the official data source for compiling R&D statistics, BOS is designed to build a better understanding of the range of business practices that impact on firm performance and so it also collects simple data on R&D expenditure.[16] The Inland Revenue Department's IR10 form also collects information on R&D expenditure within the context of a broader profit and loss statement.

Fabling (2007) compared R&D Survey and BOS responses for the sub-sample of firms that completed both forms and found that, at the micro level, BOS R&D figures are systematically lower than comparable R&D Survey figures. He reasons that this is likely to reflect an enumeration effect in the R&D Survey – that is, R&D expenditure in that survey is itemised over several pages, whereas BOS asks for a single total figure. Nevertheless, the correlation between the logs of the two measures of total R&D expenditure is quite high, being 0.872. For that reason, we believe it is valid to pool BOS and R&D survey observations taking care to adjust BOS responses for the average difference.[17] Table 2 sets out our coverage by data source and year. We pool the data because BOS is a much larger sample survey – using only 2004 and 2006 R&D Survey results we would have 4,140 observations that meet our population criteria, and with BOS included this number rises to 16,080. One consequence of pooling the data is that we change sources by year, making it difficult to isolate any effect from time-varying macro variables, such as interest rates.

Fabling (2007) also considers differences between survey responses to questions about the split between internally and externally conducted R&D. He concludes that there are serious disparities between measures of this split, probably as a result of question phrasing and response burden incentives in the R&D Survey. Overall firms in the R&D Survey are more likely to report their R&D as completely conducted in-house. If the BOS results are more accurate, this would imply some element of double-counting in BERD figures, since both funders and performers are report the same R&D as being performed by them in the R&D Survey. From a policy perspective – and to understand firm performance – we are more interested in the total investment firms make in R&D regardless of who performs the service.[18] Balancing these factors, in this paper we choose to focus on total firm-level expenditure accepting there will be an element of double-counting in the estimated aggregate we produce. From a practical perspective, it would be difficult to reconcile our BOS and R&D Survey responses to make them consistently measure the internal-external split. Insofar as business-funded research is performed outside the private-for-profit sector there will be no double-counting of investment.[19] Looking solely at total R&D reported from the survey responses that meet our population criteria, we actually calculate aggregate (weighted) R&D to be lower than that reported in official statistics. Table 3 reconciles the differences between the calculation of aggregate R&D using our population and the equivalent number reported in official statistics (ie, in Statistics New Zealand 2006). Our main differences arise because we

---

16 The overlap between the R&D Survey and BOS is quite small so that there is limited opportunity to reduce respondent load in the collection of R&D data by the sort of linking that is undertaken in this paper.

17 Specifically, for the common sample we regress ln(total R&D) from the R&D survey on ln(total R&D) from BOS and a constant. We then use predicted R&D survey values where we have a BOS response and no R&D survey response.

18 Though some research suggests that in-house R&D capability is important, particularly because such capability may increase the ability of a firm to make use of externally-generated knowledge (Cohen and Levinthal 1989).

19 Statistics New Zealand (2006) indicates that $138m of R&D funded by business is performed by government or the higher education sector, both of which are outside our population.

take a narrower population – restricting to private-for-profit and excluding non-employing firms.[20]

The IR10 would be our ideal source of R&D expenditure information, given the superior coverage that tax data provides. However, the R&D variable in the IR10 does not concord well with survey source estimates of R&D, probably because salaries and wages are separately accounted for in the IR10 form. That is, it would be impossible for firms to accurately report both their full R&D cost and their full wage bill without overstating their total expenditure. Given that the internal accounting for salaries and wages is probably superior, measurement error results in the R&D category of the IR10. The conclusion section briefly discusses how the implementation of a tax credit could greatly enhance the value of administrative data in the estimation of BERD. Still, the IR10 variable provides a reasonably good binary indicator of whether firms performed R&D in a given year and we use it in that context. The correlation between indicators of whether R&D is performed in the firm using BOS 2005 and corresponding IR10 responses is 0.669.[21]

To summarise, our dependent variables (R&D activity and R&D intensity) are based on both R&D Survey and BOS reported total R&D, where precedence is given to R&D Survey responses. Where BOS responses are used, they are adjusted to allow for differences in the survey mechanism (which are likely to cause respondents to report inconsistently across the two surveys). IR10 data is used as a consistent measure of historical R&D activity.

We now turn to our measures of the determinants of R&D investment. In discussing these variables, we follow the categorisation of the previous section (see Appendix A for a summary of the definition of the variables):

Firm size: Total employment measured as the sum of LEED rolling mean employment (RME) and an annual count of working proprietors in the firm (also sourced from LEED).

Internal financing: Following the methodology of Fabling and Grimes (2008) we use AES postal responses together with IR10 forms to construct measures of the debt-to-equity ratio (DER) and dividends to profit ratio (DTP). We choose the debt-to-equity variable because the inherent variability in returns to R&D implies firms should have higher shares of equity if they wish to undertake R&D. That is, the certain repayment schedule implied by debt is an inappropriate match to the uncertain profit stream of the R&D investment. The dividends to profit measure should signal whether firms have good future investment opportunities. Specifically, if internal financing is necessary for R&D, then firms should retain profits to fund the activity. We also include the three-year average past profitability as an additional measure of potential availability of internal funds,[22] and control for ownership type (six

---

20 The numbers in Table 3 are calculated hierarchically so that the figures total to the difference between our numbers and official data. That is, differences attributed to zero employment firms being excluded do not include zero-employment firms that are also out of scope because of business type or institutional sector.

21 This calculation uses tetrachoric correlations since these variables are binary measures of R&D (ie, do/don't do).

22 Because this variable has three lags, we can only use BAI data to construct it without losing the ability to predict R&D for earlier years. Since BAI data excludes wage and salary on the expense side, the numerator of our measure of profitability (sales-purchases) could be interpreted as being a measure of value-added. We choose to label this variable as "profitability" because of the denominator choice (sales), but will interpret the variable as representing prior performance (either profitability or productivity) more generally.

dummies) which may affect access to finance and owner/manager incentives to take on risky projects.[23]

Competition/market power: We use BAI sales to construct the firm's share of aggregate (three-digit industry) sales and an industry Herfindahl index (ind_herf) to capture how concentrated sales are within industries. We also include the square of ind_herf to allow us to test the possibility that industry concentration has an inverted-U effect on R&D activity. Unfortunately, we have no adequate control for import competition across all industries.[24]

Location: Our model does not include any location variables, although potential candidates could include the minimum distance between a New Zealand firm and the closest Crown Research Institute or university, the diversification of the local industry mix (see, eg, Maré and Timmins 2007), the share of nearby firms that conduct R&D, and/or the availability of fixed factors such as information and communications technology infrastructure (eg, broadband).

Technology characteristics: A set of 18 industry dummies are included and we also allow coefficients of some other variables to vary by a basic split of firms into high- and low-technology industries. This split is based on average industry R&D per firm so that roughly half of observations are deemed to be in each technology group. In practice, 41 percent of our firms are dubbed high-technology with industry-averaged expenditure over $13,000, constituting all of manufacturing (excluding printing, publishing and recorded media) and property and business services.

Government support: We include a dummy for whether firms are contemporaneously receiving grant money from the Foundation for Research, Science and Technology (FRST).

Human capital: We have no consistent firm-level measure of human capital over time. The LBD will soon contain some local labour market characteristics that may be useful controls. Potentially, access to individual-level data in LEED would allow some controls for the effect of flows of people between firms to be incorporated into the model

International engagement: We include a dummy for whether a firm is foreign-owned (based on IR4 and LBF data), as well as a dummy for whether firms export or not (based on zero-rated GST sales and Customs data).[25]

Persistence: As discussed earlier, we use a dummy of positive lagged IR10 R&D (prior_R&D) as a measure of persistence. We also include the ratio of intangible assets to total assets (ITA) as a proxy for the accumulated stock of knowledge (sourced from AES and IR10 data).

Macroeconomic conditions: As noted earlier, our R&D variable data source varies systematically by year, which precludes us from including general macro variables. Their potential effect would be convoluted with any remaining data inconsistencies. However, a more fruitful approach in the future might be to use average industry-level interest rates if

---

23 In cases where firms are not active in all years required for these variables to be constructed, they are set to zero and (unreported) dummies are included in the regressions for those observations. There are not sufficient observations of new entrant firms to estimate a separate model for them.

24 It is probable that the Customs data could be used to both improve the identification of market competitors and allow the construction of import competition variables. However, implementing this approach would require us to restrict our analysis to manufacturing firms.

25 Exports are not the only sales items that are zero-rated for GST purposes, so this variable is only a partial proxy for exporting behaviour. We choose to use this data in the absence of a comprehensive measure of service exports in the LBD. For a manufacturing-only analysis, the Customs data could also be used to identify the specific markets firms operate in.

these could be constructed. Additionally, exchange rate movements could be captured in import competition variables if present (as discussed above).

All financial variables are lagged one year to reduce issues of endogeneity. Despite this, causality could still be from R&D investment to the “independent” variables. For this reason, we interpret our model as suggesting correlations rather than causation.

Both IR10 and BAI data have had missing data imputed. We make use of this imputation for our bottom-up BERD calculation, but exclude imputed IR10 observations from our modeling exercise.[26] This exclusion reduces our estimation sample from 16,080 to 12,906. In the following section, we test the robustness of our model to the inclusion of imputed data.

Figure 3 shows distributions of selected variables in 2006. Many of our right-hand side variables are highly skewed, with some of this due to the way entering firms are treated. The debt-to-equity ratio (DER) is also right censored for a number of firms (because reported equity is negative). Similarly, profitability has a very long tail to the left and has been left-censored at negative one. In general, though, a large proportion of firms return a positive profit averaged over a three year period. Employment is highly concentrated at one and two staff (potentially working proprietor-only firms), with a long tail to the right (the exclusion of the top one percent of firms from this chart hides the true breadth of this distribution). Very few firms report non-zero intangible assets.

4 Results

In this section we estimate our Heckman model, appraise the effectiveness of the model and perform robustness tests on our results. Having satisfied ourselves that the model has explanatory power, we then derive a time series for aggregate BERD using predicted R&D expenditures from the model. Because of data availability, we have the advantage of producing an annual BERD series (the R&D Survey is only run every two years), and of being able to decompose the relative contribution of various factors to the estimated growth of R&D. We also decompose the investment in R&D into firm size groups by year and produce predictions of the number of private-for-profit firms investing in R&D in 2006. The section concludes by discussing the importance of deflators and recalibration over time.

Preliminary exploratory analysis suggests that the standard international result that R&D intensity does not vary with firm size does not hold in the New Zealand data. This fact leaves us with a decision to make regarding which variables should appear in our selection equation and not appear in our intensity equation. We propose that having available finances or some ability to conduct R&D should be important determinants of the decision to do R&D, but may not determine the size of the investment. For firm analysis it is often hard to justify such choices and, to a certain extent, such decisions have to be made on faith. By electing to use measures of past R&D performance and the ability to internally finance investment as our selection-only variables, we drop ITA, prior_R&D, DER and DTE from the intensity equation. We leave profitability_3yr in both selection and intensity equations because this variable may reflect more than internal financing ability (as noted in footnotes 7 and 22).

Table 4 presents our main results. Column one is the central estimate using the Heckman selection model and subsequent columns present robustness checks for that model. The first

---

26 Less than one percent of remaining observations in our estimation sample have BAI data imputed.

point of note from column one is that only 15.4 percent of observations actually do R&D (ie, the ratio of "uncensored" to total observations, 1,989/12,906).

Most of our model variables have the expected sign. In particular, in the selection equation receiving funding from FRST (FRST_funded), being an exporter, having relatively large market share (Sales_share), being in R&D intensive industries, having a balance sheet with a lower debt to equity ratio (DER), retaining profits (DTP), and having performed R&D in the prior year all raise the probability that the firm will be doing R&D. Where included, these variables also have the same positive relationship with R&D intensity, though only firms in a small subset of industries are clearly investing at higher intensities than retail trade firms (our reference group). Foreign-ownership had an ambiguous sign in our earlier discussion of theory – however, it shows up as positively related to R&D intensity in our estimates.

Industry competition has the anticipated inverted U-shaped relationship in both the selection and intensity equations (positive sign on linear term, negative sign on squared term though in the intensity equation the squared term is not significantly different from zero). However, in the selection equation, the estimated turning point of this inverted-U is at a Herfindahl index value of 0.386. Looking again at the bottom right panel of Figure 3 we can see that there are very few industries that have higher concentration than this number (containing only one percent of firms in the population), so that our estimated relationship is essentially linear over the range of observed concentrations. Firms in more concentrated industries are more likely to do R&D even after controlling for the share of the firm's own sales in that industry.

The sign of our prior performance variable (profitability_3yr) is perhaps somewhat surprising with weaker performance related to the decision to invest in R&D and the intensity of that investment too. Size, measured by total employment also has the opposite sign to expectation, though small in magnitude, in the selection equation. In our robustness tests we will focus on whether these last two estimated parameters are plausible.

Firstly, a Wald test of the independence of our selection and intensity equations cannot reject the possibility that we did not need a selection equation (p=0.199 on a test of rho, the correlation between error terms, equal to zero). That is, given our choice of "selection"-only variables it appears that a simple ordinary least squares (OLS) model would be appropriate in estimating the relationship between R&D intensity and our right-hand side variables. Columns two and three of Table 4 present OLS estimates with and without our "selection"-only variables. As expected, our OLS parameters vary only mildly from the second stage Heckman estimates presented in column one.

In columns four and five we re-estimate the two-stage model using only BOS observations.[27] We do this for two reasons – to check that our approach of using both R&D Survey and BOS data is not problematic; and to allow us to estimate a population-weighted model (in column five). Both unweighted and weighted BOS-based models present consistent coefficients with the base model (in terms of sign and significance). We would expect some variation in the results for the intensity equation simply because we have lost almost half of our R&D performers. That is, while the R&D Survey contributes a small proportion of total observations, it contributes an equal share of observations of non-zero R&D expenditure. The main changes in estimated coefficients from the base model occur for profitability_3yr, employment, industry concentration and the sole proprietorship and partnership dummies. The BOS-only models suggest the estimated coefficient of the prior performance variable is

[27] The BOS population does not include firms in ANZSIC Division Q (Personal and Other Services), so the ind_Q dummy is dropped for these columns.

somewhat fragile. Specifically, while still negative in sign, we can no longer reject the possibility that profitability_3yr is unrelated to R&D selection or intensity (in the weighted model). The changes in coefficients on log employment and the sole proprietor dummy may reflect changes in the population arising from only using BOS observations. BOS has an employment cut-off of six rolling mean employees, whereas the R&D Survey includes firms with much lower employment. For sole proprietors and partnerships, it is quite plausible that those owner-operated firms with six employees or more are very different from employer-only observations. For the employment variable, it is possible that very low employment firms are predominantly driving the negative relationship between R&D intensity and firm size (that is, the very small firms in our full dataset are very R&D intensive). This may be the case because the R&D Survey targets likely R&D performers so that R&D Survey observations with very low employment may not be "representative" small firms when it comes to doing R&D. Alternatively, there may be measurement error in the employment variable. Including this variable as the denominator on the left-hand side will then generate a strong estimated negative relationship with the same variable on the right-hand side. Our results in columns four and five are, therefore, consistent with that measurement error declining with firm size (as is likely to be the case).

Finally, in column six, we include observations with imputed IR10 data. We do this to check that our results aren't biased because missing IR10 data is somehow correlated with doing R&D. The results of this test suggest we do not have a problem in this regard.

The other robustness test – suggested earlier in the paper – was to check whether investment decisions in high and low-technology industries have different drivers. This hypothesis was tested by interacting firm variables with a dummy for being in a high-technology industry. The results are not reported in Table 4 because the only significant difference between high- and low-technology firms in the intensity equation arises for profitability_3yr. In particular, the negative relationship between prior performance and R&D intensity is exclusively driven by high-technology firms. The coefficient for high-technology firms is -0.691, significant at the one percent level, compared to 0.204 for the low-technology firms, not significantly different from zero at the 10 percent level.

Overall, our robustness tests suggest that our model may not require a Heckman two-stage specification, but that most estimated relationships are consistent with theory, prior evidence, and are robust to variations in the data used to estimate the parameters. Given this conclusion, we proceed to estimate our aggregate BERD series and explain its evolution. We do this by aggregating the expected value of R&D performed at each firm. This expected value is calculated as the predicted R&D expenditure conditional on the decision to do R&D (ie, a prediction from the second stage equation) multiplied by the predicted probability that the firm chooses to perform R&D (from the first stage equation). Thus our model estimates a non-zero expected R&D expenditure for every firm in the population.

Table 5 presents our estimate of aggregate BERD using the model in column 1 of Table 4. We estimate aggregate BERD in 2006 to be approximately 65 percent higher than the weighted R&D Survey observations in our population would suggest.[28] This number is substantially larger but not necessarily implausible. Our difference comes mainly from the estimated R&D activity of firms with low employment. Table 5 decomposes our modeled aggregate into the contribution of firms that have less than ten employees, those with 100 or more employees, and an intermediate size group. The smallest firms contribute almost three quarters of aggregate expenditure. This is not because the small size group is expected to do

[28] Our estimate also sits outside the 95 percent confidence interval for the weighted R&D Survey result.

more R&D, on average, but rather because they constitute 94 percent of all firms in the population. The R&D survey design allocates most firms in this size category to an "unlikely to be doing R&D" sampling strata that has a relatively low sampling rate, or excludes them from the population. The large employment group make up only around one third of one percent of firms but account for over nine percent of aggregate modeled BERD.[29]

Aggregate growth in R&D is estimated to be a compounding 2.4 percent per annum. Figure 4 shows the evolution of a selected set of the modeled correlates of R&D expenditure. Each variable mean is indexed to 100 in 2001 and the legend is ordered based on each index's relative value in 2006. From this simple breakdown, it appears that increasing numbers of exporters, more firms developing R&D experience, increased levels of foreign-ownership, and higher average retained earnings have contributed to the overall increase in predicted R&D. Another key contributor is evident from Table 1, where it can be seen that there has been a steady decrease in the share of firms that are in primary production. Our industry dummies, and cross-country evidence such as that presented in Crawford *et al.* (2007), suggest that these firms are relatively weak R&D performers.[30]

While our aggregate series is very stable over time, the micro model predicts that many firms vary their investment year-on-year. Figure 5 shows the distribution of the year-on-year log difference in expected R&D. Many firms are predicted to make small changes to their R&D investment with growth centred roughly at zero, and not particularly skewed. The clustering around zero is partly driven by the fact that some important model variables are unlikely to change over time or do so only infrequently, such as industry or ownership structure.

An important question is whether the model needs to be recalibrated over time. Recalibration would be necessary if R&D intensity changes over time within firms that maintain the same controlled-for characteristics. For example, if R&D intensity increases within firms with otherwise the same employment, etc, then the model will underestimate aggregate BERD growth. It is not clear that this should be the case for the last decade. Any such effect would depend on some inadequacy in our controls for determining R&D expenditure, or some instability in the model coefficients over time.[31] Given the absence of sample survey data for years prior to 2004, it would be very difficult to empirically test the stability of the model. Specifically, the selection equation would become very difficult to estimate because the observations increasingly come only from firms that are expected to be R&D performers. Similarly, estimating within-firm intensity changes based on a panel approach (ie, by looking only at growth in intensity from firms that appear in consecutive years) is likely to reintroduce the problem that our method is trying to avoid – that of biased sample selection. For example, in earlier years firms that are no longer expected to be doing R&D would be dropped from subsequent population lists, so that the panel estimation of the persistence effect becomes meaningless. Further, historically, addition to the sample could be determined

---

29 The model assigns the smallest firms an average probability of being R&D performers of 9.0 percent, the intermediate size group a probability of 10.0 percent, and the largest group a probability of 16.1 percent. That is, large firms are far more likely to be R&D performers, despite the estimated coefficient on log employment in the selection equation being negative.

30 Some of the decline in the number of agriculture firms may be due to changing land use from farming to residential property. We would not expect to see R&D to be growing in this case, since property management companies tend not to be intensive R&D performing firms either. However, our population does not exhibit particularly large increases in property and business services firms of this sort, perhaps because these units would largely appear on the LBD as non-employing.

31 The model may not be stable if, for example, the funding structure of BERD or other parts of the national innovation system have changed over the decade (eg, if public research institutions began producing more business-relevant research due to new incentives).

by one of our independent variables and may be related to rapid growth in R&D. For example, new FRST funding recipients were automatically added to the survey list. These recognised problems with purely list-based approaches are part of the reason Statistics New Zealand has moved to a sample methodology more closely aligned with other official business surveys.

Our estimated model has good explanatory power and is largely consistent with prior research and theoretical expectations so that we have little evidence to suggest recalibration is necessary. Therefore, on the assumption that our model is a good representation of the R&D investment decision, and that any recalibration might be argued to be arbitrary, we do not recalibrate our aggregate series.

However, there is still the question of how we should appropriately deflate the R&D expenditure series. All of our independent financial variables are expressed in ratios that might reasonably be expected to have the same deflator in the denominator and numerator. Our R&D intensity variable, on the other hand, is not deflated. This may present an issue in estimation, though that might be considered relatively minor over the three year estimation period. However, the problem becomes more material over a six years period. Generally R&D expenditure has been implicitly deflated by the GDP deflator by dividing nominal BERD by nominal GDP. However, recent research suggests that R&D-specific deflators make a substantial difference to the interpretation of cross-country differences in R&D intensity (Dougherty *et al.* 2007). In particular, once relative labour costs are factored in, cross-country differences in R&D activity are much smaller. As we noted earlier R&D is highly skills-intensive and relative demand (and therefore wages) for skilled labour has been increasing steadily, so that it is not clear that growth in BERD has even matched growth in input costs. The simple average of Statistics New Zealand’s Labour Cost Index (LCI) for professionals, and technicians and associate professionals has grown by 2.5 percent per annum between March 2001 and March 2006. On the basis of the LCI change, we could conclude that our model predicts the level of real R&D to be static over 2001-2006.

5 Conclusions and policy implications

In this paper we have investigated the feasibility of using a microeconometric model to bridge methodological changes in survey design and provide consistent estimates of a macro aggregate over time. We reach the conclusion that the method is feasible, and that it provides a clear indication that business expenditure on R&D has been underreported historically. Specifically, we predict BERD to be 65 percent higher than official statistics (for our chosen sub-population) in 2006, with this gap rising rapidly back through time. In real terms, our model predicts that BERD has been static over the estimation period, though from a much larger base than previously thought. Much of the “additional” R&D predicted by the model is generated by the extremely large number of smaller firms that are generally treated as not being “likely” R&D performers. These findings have implications for research, policy and survey collection. Each of these is discussed in turn.

Firstly, for sound theoretical reasons, researchers tend to accumulate R&D investments into stocks described as “knowledge capital”. A recent example of this approach is given in Hall and Scobie (2006), which looks at the productivity of R&D in the agricultural sector and uses a perpetual inventory method to accumulate R&D expenditure into a stock. If historical R&D has been underestimated over time, the stock has been underestimated, and the returns on R&D have probably been over-estimated. Even in the absence of stock accumulation, cross-country panel estimates (which rely implicitly on the same variable being measured across different jurisdictions) including New Zealand could produce inaccurate results. Crawford *et al.* (2007) find that New Zealand's R&D per capita is not unusual (by OECD standards) after fundamental characteristics of the economy are accounted for. It is possible that New Zealand should actually be regarded as a positive outlier, given our predicted values of BERD.

From a policy perspective this paper suggests that New Zealand's historical innovative performance is better than previously supposed. However, it also suggests that we have not recently observed a "step change" to a new higher level of BERD, since growth in BERD is almost certainly slower in nominal terms after accounting for survey design changes. The introduction of the tax credit is forecast to stimulate additional R&D in the business sector (IRD and Treasury 2007). Given that we estimate a higher existing level of BERD, our results potentially imply that the scheme costs are higher than previously estimated. Specifically, the policy was costed using 2004 R&D Survey figures, which estimate (using our modified dataset) roughly 1,700 firms doing some R&D with 1,300 currently meeting the $20,000 threshold for credit eligibility. Using the expected R&D spend, we predict there are 8,204 firms currently eligible for the credit (that is, with expected R&D greater than the threshold). However, caution should be taken in assuming that these larger estimated numbers of eligible R&D performers necessarily imply a more expensive tax credit scheme than forecast. In particular, the costing of the scheme also assumed a stimulatory effect – based off research in other countries – of new R&D caused by the introduction of the tax credit (IRD and Treasury 2007). Those foreign growth rates may in part capture the foreign tax credit leading to identification of R&D performers that were already conducting R&D, but not captured in pre-credit statistics. That is, the IRD & Treasury costing may well be a perfectly good predictor of the scheme cost if the overseas-derived estimate of the growth effect of introducing a tax credit has been estimated off a similarly underestimated measure of BERD.

Finally, the introduction of the tax credit is likely to seriously improve the administrative data available to Statistics New Zealand in the estimation of BERD. While the policy rationale, overseas experience, and our results all point toward a step change in the reported level of R&D, it is unclear how much of that will be because firms:

- invest more in R&D because they are now compensated for externalities;
- start reporting credit-compliant activity that is not "really" R&D; and
- are incentivised to self-identify and account for R&D activity they are already doing.

This paper suggests that, as with the introduction of the survey sample methodology, the arrival of high-quality administrative R&D data is likely to result in the discovery of lots of existing R&D performers (ie, firms that fit under the third bullet point).[32] It will be a challenging task for subsequent evaluations to separate this measurement effect from the policy stimulus effect. Statisticians and policymakers in New Zealand are well aware of these challenges and are working hard to make sure the data collected is up to the task of providing a robust basis for future policy advice.

---

32 The survey stratification of the R&D Survey (and subsequent estimated counts of R&D performers) relies on the ability to identify a very large pool of firms that are expected not to be R&D performers. Our conclusion relies on an assessment that this pool probably contains a significant number of R&D performers. For example, if we use our model and aggregate predicted R&D expenditure (conditional on doing R&D) for those firms that have a predicted probability of doing R&D of at least 50 percent, then we estimate aggregate BERD as only $227m in 2006. In other words, the assessment of who is a likely R&D performer is a critical determinant of the resulting estimate of aggregate BERD.

References

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith & Peter Howitt (2005) “Competition and innovation: An inverted-U relationship”, Quarterly Journal of Economics 120(2), 701-728

Becker, Bettina & Nigel Pain (2008) “What determines industrial R&D expenditure in the UK?”, The Manchester School 76(1), 66-87

Cohen, Wesley & Daniel Levinthal (1989) “Innovation and learning: The two faces of R&D”, The Economic Journal 99(397), 569-596

Crawford, Ron, Richard Fabling, Arthur Grimes & Nick Bonner (2007) “National R&D and patenting: Is New Zealand an outlier?”, New Zealand Economic Papers 41(1), 69-90

David, Paul, Bronwyn Hall and Andrew Toole (2000) “Is public R&D a complement or substitute for private R&D? A review of the econometric evidence”, Research Policy 29, 497-529

Davis, Nick (2006) “Business R&D, innovation and economic growth: An evidence-based synthesis of the policy issues”, Ministry of Economic Development Occasional Paper 06/08

Dougherty, Sean, Robert Inklaar Robert McGuckin & Bart van Ark (2007) “International comparisons of R&D expenditure: Does an R&D PPP make a difference?”, NBER Working Paper 12829

Fabling, Richard (2007) “Comparison of reported R&D from BOS’06 and R&D’06”, Statistics New Zealand mimeo

Fabling, Richard & Arthur Grimes (2008) “Over the hedge or under it? Exporters’ optimal and selective hedging choices”, paper presented at NZAE’08, Wellington, New Zealand

Fabling, Richard, Arthur Grimes, Lynda Sanderson and Philip Stevens (2008) “Some rise by sin, and some by virtue fall: Firm dynamics, market structure and performance”, Ministry of Economic Development Occasional Paper 08/01

Furman, Jeffrey, Michael Porter & Scott Stern (2002) “The determinants of national innovative capacity”, Research Policy 31, 899–933

Hall, Julia & Grant Scobie (2006) “The role of R&D in productivity growth: The case of agriculture in New Zealand: 1927 to 2001”, Treasury Working Paper 06/01

Inland Revenue Department & New Zealand Treasury (2007) “Tax policy report: BTR tax credit design package: cap and associated measures”, report to the Minister of Finance and Minister of Revenue PAD2007/22 T2007/112

Jaffe, Adam, Manuel Trajtenberg & Rebecca Henderson (1993) “Geographic localization of knowledge spillovers as evidenced by patent citations”, The Quarterly Journal of Economics 108(3), 577-598

Johnson, Robin, Weshah Razzak & Steven Stillman (2007) "Has New Zealand benefited from its investments in research & development?", Applied Economics 39(19), 2425-2440

Maré, David & Jason Timmins (2007) “Geographic concentration and firm productivity”, Ministry of Economic Development Occasional Paper 07/01

Mazoyer, Pamela (1999) “Analysis of R&D structure and intensities”, Ministry of Research, Science and Technology Research Paper

MoRST (2006a) “Research and development in New Zealand: A decade in review”, Ministry of Research, Science and Technology: Wellington

MoRST (2006b) “The development of a time series of business expenditure R&D statistics”, Ministry of Research, Science and Technology: Wellington

OECD (2006) “OECD science, technology and industry outlook 2006”, OECD: Paris

OECD (2007) “OECD reviews of innovation policy: New Zealand”, OECD: Paris

Romer, Paul (1990) “Endogenous technological change," Journal of Political Economy 98(5)

Smith, Keith (2006) "Public policy framework for the New Zealand innovation system", MED Occasional Paper 06/06

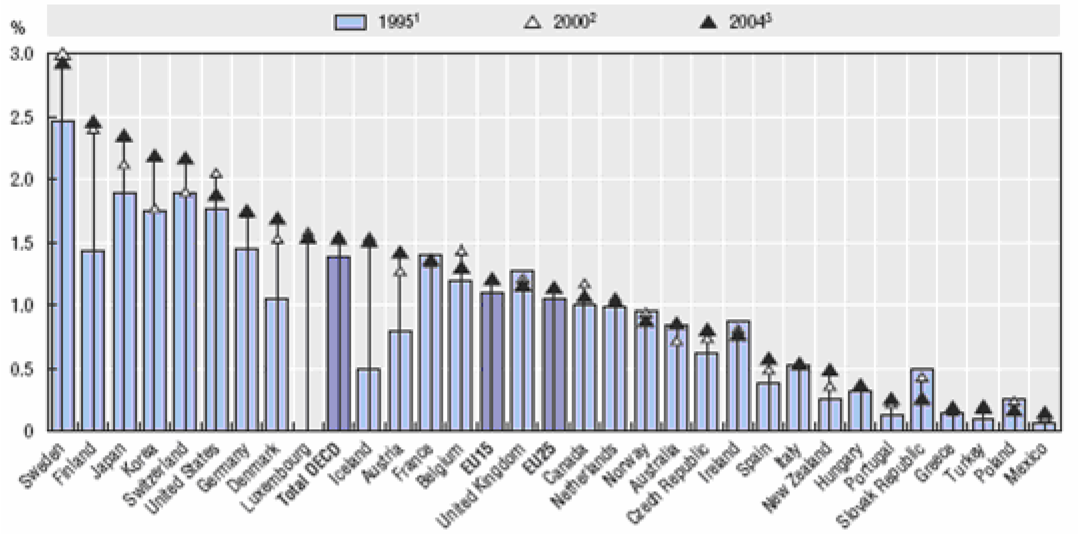
Statistics New Zealand (2002, 2004, 2006) "Research and development in New Zealand", Statistics New Zealand: Wellington

Statistics New Zealand (2007) "Prototype outputs from the Longitudinal Business Database", Statistics New Zealand: Wellington

Symeonidis, George (1996) "Innovation, firm size and market structure: Schumpeterian hypotheses and some new themes", OECD Economics Department Working Paper 161

FIGURES & TABLES

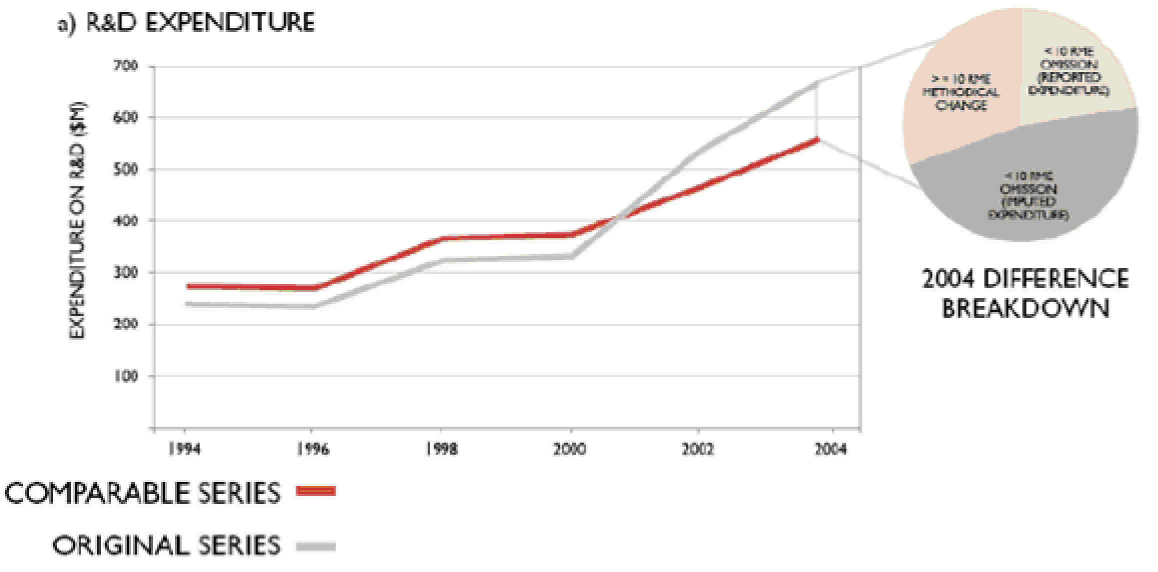
Figure 1 – BERD as a percentage of GDP by country, 1995, 2000 and 2004



1. 1993 for Austria, 1996 for Switzerland.
2. 1998 instead of 2000 for Austria; 1999 for Denmark, Norway, New Zealand and Sweden.
3. 2002 for Austria and Turkey; 2003 for Australia, Greece, Iceland, Mexico, New Zealand, Portugal and Sweden.

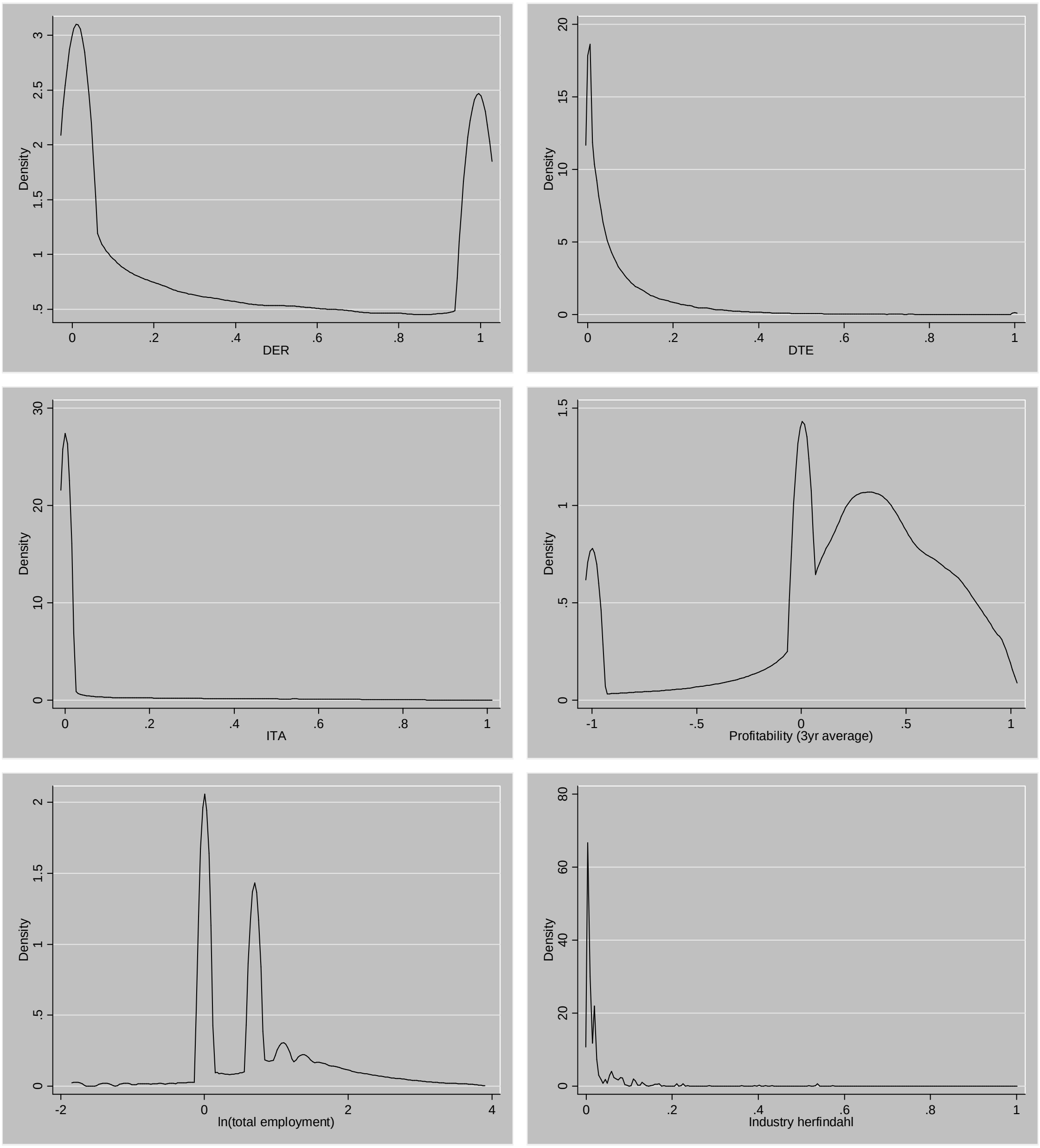
Source: OECD (2006)

Figure 2 – MoRST revisions to BERD measure



Source: MoRST (2006b)

Figure 3 – Distribution of selected right-hand side variables in 2006



Note: The kernel density plot for ln(total employment) excludes the top and bottom one percent of observations for confidentiality reasons.

Figure 4 – Average value of selected right-hand side variables over time

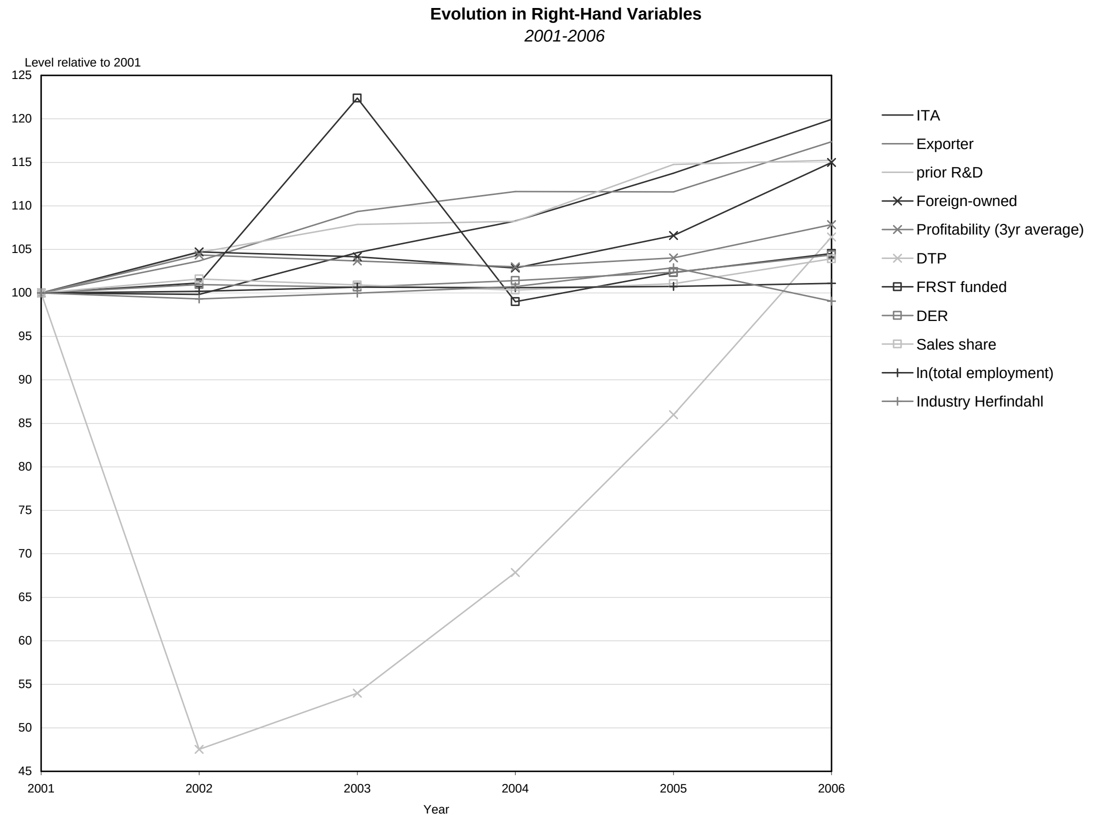
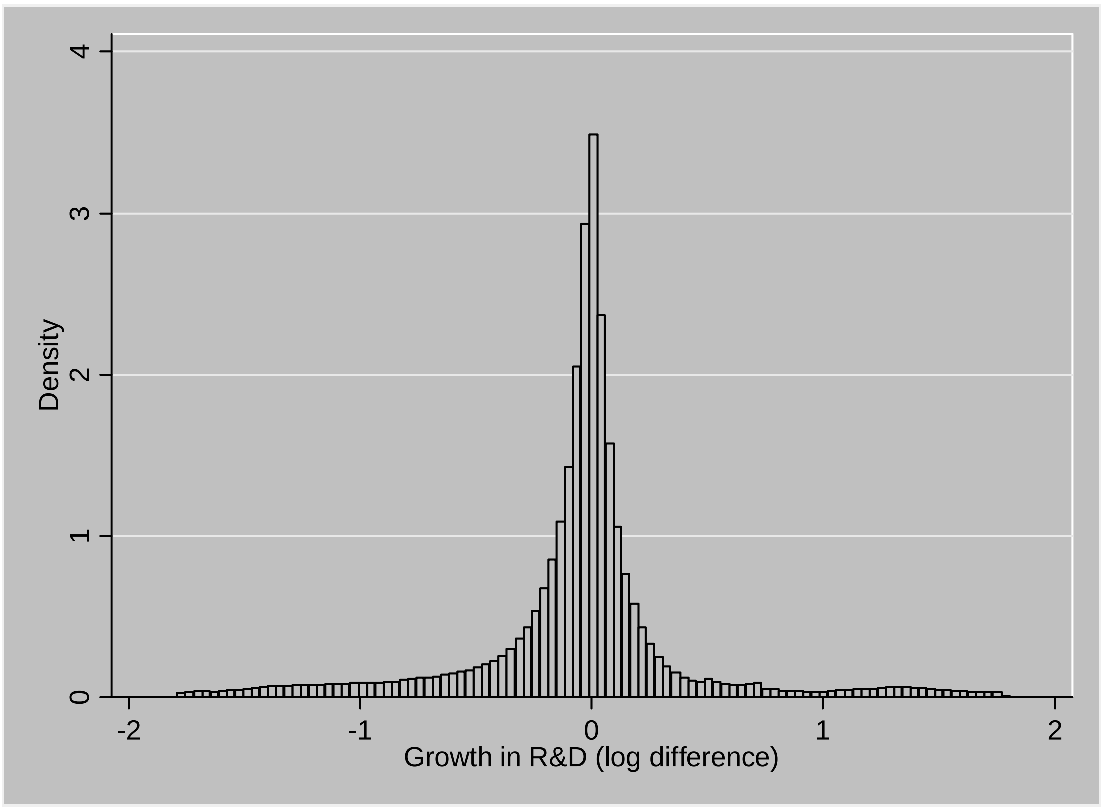


Figure 5 – Distribution of annual growth in expected R&D, all years pooled



Note: The top and bottom one percent of observations have been dropped for presentation purposes.

Table 1

Population Counts

By industry

2001–2006

	Year					
	2001	2002	2003	2004	2005	2006
	Number of enterprises					
Industry						
Agriculture, forestry and fishing	81,756	79,659	76,422	73,836	70,242	65,349
Mining and quarrying	324	324	297	297	282	300
Manufacturing	23,412	23,250	23,322	23,256	23,013	22,440
Electricity, gas and water supply	60	63	81	81	93	102
Construction	40,359	39,738	40,383	42,048	43,563	44,064
Wholesale trade	15,714	15,588	15,429	15,489	15,306	14,940
Retail trade	40,038	39,321	39,411	39,603	39,459	38,562
Accommodation, cafes and restaurants	11,220	11,580	12,081	12,456	12,681	12,630
Transport and storage	11,967	11,736	11,790	11,892	12,057	11,862
Communication services	3,657	3,564	3,510	3,546	3,468	3,210
Finance and insurance	4,323	4,560	4,770	4,908	5,154	5,145
Property and business services	70,071	69,438	72,894	74,709	75,294	74,052
Education	2,088	2,226	2,370	2,574	2,610	2,610
Health and community services	13,143	13,155	13,281	13,557	13,854	13,761
Cultural and recreational services	7,077	7,161	7,392	7,611	7,761	7,506
Personal and other services	9,924	10,014	10,239	10,488	10,632	10,488
Total	**335,136**	**331,377**	**333,669**	**336,357**	**335,469**	**327,021**
Entries and exits(1)						
Entries (start-up enterprises)	44,883	44,673	49,077	50,676	50,148	49,683
Exits (ceasing enterprises)	44,025	48,432	46,785	47,988	51,036	58,131

(1) Entries and exits include firms shifting between employing and non-employing.

Note: All counts were randomly rounded to base 3 to protect confidentiality, so actual figures may differ from those stated. Due to rounding, some figures may not add to stated total.

Table 2

R&D Expenditure Data Source

By survey

2004–2006

	Year		
	2004	2005	2006
	Number of enterprises		
Business Operations Survey (BOS)		6,879	5,784
Less those also in R&D Survey(1)			723
Total BOS		6,879	5,061
Research and Development (R&D) Survey	1,812		2,328
Overall total	**1,812**	**6,879**	**7,389**

(1) R&D Survey responses were used where firms responded to both surveys.

Note: All counts were randomly rounded to base 3 to protect confidentiality, so actual figures may differ from those stated. Due to rounding, some figures may not add to stated total.

Table 3

Reconciliation Between R&D Survey Sample and Official Statistic[1]

2006

	$(million)
Research and Development (R&D) Survey[2]	**931.5**
Model - using weighted R&D Survey total expenditure	**734.3**
Reason for difference	
Out of scope for population	
Business type	94.7
Not active or institutional sector	34.8
Zero total employment	64.8
Other	2.9
Total difference	**197.2**

(1) BERD figures refer to in-house and funded activities.

(2) The report *Research and Development in New Zealand 2006* only publishes the in-house portion of BERD ($763.3million).

Table 4 – Heckman model results

	(1)		(2)	(3)	(4)		(5)		(6)	
Model type	Heckman		OLS	OLS	Heckman		Heckman		Heckman	
Dependent R&D variable	Intensity	Selection	Intensity	Intensity	Intensity	Selection	Intensity	Selection	Intensity	Selection
profitability_3yr	-0.400***	-0.268***	-0.445***	-0.396***	-0.241	-0.136**	-0.392	-0.002	-0.465***	-0.291***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.323]	[0.046]	[0.341]	[0.987]	[0.000]	[0.000]
FRST_funded	0.733***	1.153***	0.869***	0.836***	1.052***	0.922***	1.516***	1.050***	0.627***	1.165***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.009]	[0.000]	[0.000]	[0.000]
Foreign_owned	0.545***	-0.042	0.534***	0.520***	0.719***	-0.117**	0.800***	-0.097	0.601***	-0.011
	[0.000]	[0.378]	[0.000]	[0.000]	[0.000]	[0.041]	[0.009]	[0.313]	[0.000]	[0.803]
Exporter	0.301**	0.574***	0.391***	0.382***	0.363*	0.497***	1.075*	0.549***	0.306***	0.574***
	[0.012]	[0.000]	[0.000]	[0.000]	[0.060]	[0.000]	[0.081]	[0.000]	[0.006]	[0.000]
ln(total_employment)	-0.672***	-0.061***	-0.680***	-0.679***	-0.465***	0.099***	-0.302***	0.070**	-0.696***	-0.075***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.046]	[0.000]	[0.000]
Sales_share	1.455***	0.896***	1.538***	1.671***	1.378**	0.529**	1.321*	0.677**	1.396***	0.926***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.012]	[0.012]	[0.092]	[0.016]	[0.000]	[0.000]
Industry_Herf	1.520**	1.419***	1.686**	1.645**	1.768	0.485	-2.519	1.193	2.272***	1.305***
	[0.040]	[0.000]	[0.023]	[0.027]	[0.159]	[0.183]	[0.305]	[0.106]	[0.001]	[0.000]
Industry_Herf^2	-1.403	-1.838***	-1.629	-1.638	-1.741	-0.525	4.65	-1.203	-2.452**	-1.725***
	[0.282]	[0.000]	[0.216]	[0.211]	[0.420]	[0.316]	[0.164]	[0.237]	[0.040]	[0.000]
Sole_proprietor	-0.122	-0.161	-0.15	-0.163	0.982***	-0.638*	1.399***	-0.484	-0.293	0.018
	[0.704]	[0.326]	[0.639]	[0.615]	[0.001]	[0.090]	[0.002]	[0.236]	[0.217]	[0.881]
Partnership	-0.624***	-0.159**	-0.663***	-0.672***	-0.034	-0.135	-0.151	-0.069	-0.760***	-0.165**
	[0.000]	[0.044]	[0.000]	[0.000]	[0.918]	[0.243]	[0.706]	[0.671]	[0.000]	[0.013]
Co_op	1.791***	0.688***	1.870***	1.937***	2.580***	0.333	1.593***	0.296	1.836***	0.653***
	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]	[0.236]	[0.004]	[0.232]	[0.000]	[0.005]
Joint_venture_consortia	2.817***	-0.427	2.749***	2.512***	3.441***	-0.129	2.902***	-0.678**	2.834***	-0.604*
	[0.000]	[0.220]	[0.000]	[0.000]	[0.000]	[0.707]	[0.000]	[0.035]	[0.000]	[0.075]
Overseas_branch	-0.084	-0.261	-0.144	-0.087	0.041	-0.285	-0.648	-0.601**	0.112	-0.16
	[0.805]	[0.236]	[0.669]	[0.804]	[0.930]	[0.334]	[0.437]	[0.022]	[0.739]	[0.328]
SOE	-0.267	0.201	-0.219	-0.229	0.173	0.047	0.158	0.15	-0.061	0.313
	[0.478]	[0.441]	[0.565]	[0.552]	[0.814]	[0.875]	[0.847]	[0.603]	[0.865]	[0.214]
ind_A_B	0.918	0.904***	1.069	1.056	1.268	0.817***	3.158	0.626***	0.728	0.857***
	[0.189]	[0.000]	[0.126]	[0.124]	[0.141]	[0.000]	[0.106]	[0.007]	[0.169]	[0.000]
ind_C21	0.869	0.999***	1.05	1.038	0.907	1.021***	2.571	0.613**	0.816	0.964***
	[0.212]	[0.000]	[0.130]	[0.129]	[0.288]	[0.000]	[0.204]	[0.022]	[0.125]	[0.000]
ind_C22_C23_C24	0.693	0.719***	0.832	0.794	0.663	0.741***	2.411	0.381	0.642	0.646***
	[0.316]	[0.000]	[0.230]	[0.244]	[0.425]	[0.000]	[0.196]	[0.120]	[0.221]	[0.000]
ind_C25	1.271*	1.100***	1.465**	1.426**	1.116	1.085***	2.716	0.336	1.116**	1.075***
	[0.068]	[0.000]	[0.035]	[0.036]	[0.192]	[0.000]	[0.147]	[0.211]	[0.038]	[0.000]
ind_C26	0.55	0.911***	0.717	0.644	0.285	1.050***	2.286	0.559*	0.425	0.809***
	[0.449]	[0.000]	[0.324]	[0.367]	[0.748]	[0.000]	[0.265]	[0.065]	[0.447]	[0.000]
ind_C27_C28_C29	1.683**	1.081***	1.873***	1.827***	1.659**	1.122***	2.857	0.682***	1.542***	0.993***
	[0.015]	[0.000]	[0.006]	[0.007]	[0.048]	[0.000]	[0.137]	[0.007]	[0.003]	[0.000]
ind_D	2.848**	0.057	2.845**	2.931**	5.379***	-0.006	6.398***	-0.421	2.640*	-0.065
	[0.037]	[0.887]	[0.040]	[0.043]	[0.000]	[0.991]	[0.000]	[0.451]	[0.052]	[0.868]
ind_E	0.755	0.553***	0.846	0.855	0.701	0.505***	1.426	0.282	0.735	0.436***
	[0.299]	[0.000]	[0.249]	[0.238]	[0.419]	[0.001]	[0.410]	[0.312]	[0.190]	[0.001]
ind_F	1.648**	0.465***	1.750**	1.731**	1.827**	0.398***	2.495	0.051	1.521***	0.407***
	[0.017]	[0.000]	[0.012]	[0.012]	[0.029]	[0.004]	[0.136]	[0.844]	[0.004]	[0.000]
ind_H	0.444	0.572***	0.529	0.496	0.734	0.574***	2.598	0.671**	0.209	0.493***
	[0.562]	[0.001]	[0.493]	[0.518]	[0.427]	[0.001]	[0.219]	[0.030]	[0.756]	[0.001]
ind_I	0.953	0.028	0.951	0.971	1.214	0.111	3.258*	-0.014	0.940*	-0.054
	[0.189]	[0.871]	[0.195]	[0.181]	[0.165]	[0.518]	[0.057]	[0.964]	[0.091]	[0.721]
ind_J	1.349	0.475**	1.436*	1.472*	0.999	0.477**	1.63	-0.059	0.995	0.323**
	[0.101]	[0.010]	[0.083]	[0.074]	[0.464]	[0.023]	[0.411]	[0.850]	[0.147]	[0.050]
ind_K	1.699**	0.323**	1.752**	1.789**	2.226**	0.449***	3.817**	0.075	1.658***	0.215
	[0.019]	[0.030]	[0.016]	[0.014]	[0.014]	[0.004]	[0.033]	[0.777]	[0.003]	[0.109]
ind_L77_L783	2.532***	0.994***	2.690***	2.684***	2.872***	0.922***	4.250**	0.650**	2.424***	0.932***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.024]	[0.011]	[0.000]	[0.000]
ind_L78_notL783	2.581***	0.834***	2.725***	2.700***	1.830**	0.383***	2.736	0.308	2.414***	0.813***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.033]	[0.006]	[0.142]	[0.272]	[0.000]	[0.000]
ind_N	1.165	0.837***	1.318*	1.257*	1.512*	0.933***	3.343*	0.595**	1.099*	0.706***
	[0.110]	[0.000]	[0.070]	[0.079]	[0.091]	[0.000]	[0.094]	[0.025]	[0.052]	[0.000]
ind_O_P	1.313*	0.476***	1.396*	1.356*	1.704*	0.376**	1.951	0.14	1.117**	0.399***
	[0.068]	[0.001]	[0.054]	[0.057]	[0.056]	[0.013]	[0.289]	[0.623]	[0.050]	[0.001]
ind_Q	2.056*	1.256***	2.253**	2.210**					1.964**	0.902**
	[0.058]	[0.002]	[0.037]	[0.042]					[0.042]	[0.016]
DER		-0.110**		-0.099		-0.073		-0.11		-0.112***
		[0.014]		[0.342]		[0.194]		[0.401]		[0.004]
DTP		-0.107*		-0.241*		-0.038		0.038		-0.092*
		[0.064]		[0.099]		[0.595]		[0.867]		[0.063]
ITA		0.022		0.084		0.252		0.001		-0.011
		[0.868]		[0.778]		[0.132]		[0.998]		[0.923]
prior_R&D		0.815***		0.125		0.840***		0.910***		0.747***
		[0.000]		[0.124]		[0.000]		[0.000]		[0.000]
Constant	8.640***	-1.967***	8.206***	8.241***	7.011***	-2.628***	3.901	-2.357***	8.896***	-1.828***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.209]	[0.000]	[0.000]	[0.000]
Observations	1989	12906	1989	1989	1047	10272	1047	10272	2493	16080
R-squared			0.454	0.458						

Robust p values in brackets (* significant at 10%; ** significant at 5%; *** significant at 1%)
Counts of observations randomly rounded to base 3
Column (1) is our preferred model
Columns (2) & (3) are OLS estimates of the R&D intensity equation without and with the selection variables (respectively)
Columns (4) and (5) restrict the sample to unadjusted BOS observations of R&D expenditure without and with sample weights (respectively)
Column (6) is the preferred model including observations with imputed IR10 variables
Controls for not being live in all of the last three years, and for not being live in the last year included but not reported

Table 5

Modeled expected BERD

By total employment

2001–2006

	Year					
	2001	2002	2003	2004	2005	2006
	$(million)					
Firm size						
Less than ten total employment	814.4	821.1	880.6	900.7	909.2	896.0
Between 10 and 99.99 total employment	169.0	184.8	190.4	188.7	197.2	203.6
100 or more total employment	92.7	102.4	116.8	106.9	116.4	114.6
Total	**1,076.1**	**1,108.2**	**1,187.8**	**1,196.3**	**1,222.7**	**1,214.2**

APPENDIX A – DATA DEFINITIONS

Variable	Description	Source
R&D_intensity	ln(Total expenditure on R&D/total employment). BOS responses adjusted to reflect levels difference in reporting between BOS & R&D Survey	R&D ‘04, ‘06 BOS’05, ‘06
ln(total employment)	Log of total employment, defined as rolling mean employment plus an annual count of working proprietors	LEED
prior_R&D	Binary indicator of R&D expenditure in prior year	IR10
foreign_owned	Binary indicator of foreign ownership	LBF, IR4
exporter	Binary exporter indicator	BAI, Customs
profitability_3yr	Average profit (sales less purchases) over sales for the three prior years	BAI
DER	Debt over (debt+equity) ratio for prior year	AES, IR10
ITA	Intangibles to total assets ratio for prior year	AES, IR10
DTP	Dividends to profit ratio for prior year	AES, IR10
sales_share	Share of firm sales in aggregate 3-digit industry sales for prior year	BAI
industry_herf industry_herf^2	Industry Herfindahl (sum of squared sales shares) for three-digit industry in prior year, and its square	BAI
FRST_funded	Binary of receipt of FRST funding in current year	FRST
business type	Business type binaries (sole proprietor, partnership, co-op company, joint venture/consortia, branches of overseas companies, state-owned enterprise) *Relative to limited liability company*	LBF
ind_*	Eighteen binary industry variables using the following groupings of Australian and New Zealand Standard Industrial Classifications: • A (agriculture, forestry and fishing) & B (mining and quarrying) • C21 (food, beverage and tobacco) • C22 (textile, clothing, footwear and leather manufacturing), C23 (wood and paper product manufacturing) & C24 (printing, publishing and recorded media) • C25 (petroleum, coal, chemical and associated product manufacturing) • C26 (non-metallic mineral product manufacturing) • C27 (metal product manufacturing), C28 (machinery and equipment manufacturing) & C29 (other manufacturing) • D (electricity, gas and water supply) • E (construction) • F (wholesale trade) • H (accommodation, cafes and restaurants) • I (transport and storage) • J (communication services) • K (finance and insurance) • L77 (property services) and L783 (computer services) • L78 (business services) excluding L783 • N (education) • (health and community services) & P (cultural and recreational services) • Q (personal and other services) *Relative to G (Retail Trade)*	LBF