Some rise by sin, and some by virtue fall: Firm dynamics, market structure and performance

Richard Fabling, Arthur Grimes, Lynda Sanderson & Philip Stevens

Ministry of Economic Development

Occasional Paper 08/01

May 2008

ISBN: 978-0-478-31656-8 (PDF)
ISBN: 978-0-478-31664-3 (Online)

Ministry of Economic Development
Manatū Ōhanga

Ministry of Economic Development Occasional Paper 08/01

Some rise by sin, and some by virtue fall: Firm dynamics, market structure and performance

Date: May 2008

Authors: Richard Fabling (Reserve Bank of New Zealand)
Arthur Grimes (Motu Economic & Public Policy Research)
Lynda Sanderson (Ministry of Economic Development)
Philip Stevens (Ministry of Economic Development)

Acknowledgements

This paper is largely based on earlier work by Fabling (2007a) conducted while on secondment to Statistics New Zealand. Richard thanks Eileen Basher, Claire Powell, and Steve Walshe (all from Statistics New Zealand) for valuable feedback on the production of that earlier work.

Contact: Occasionalpapers@med.govt.nz

Disclaimer

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The research was funded by the Ministry of Economic Development and supported by Statistics New Zealand as part of the Improved Business Understanding via Longitudinal Database Development project (IBULDD). The results of this study are based in part on tax data supplied by Inland Revenue to Statistics New Zealand under the Tax Administration Act. This tax data must be used only for statistical purposes, and no individual information is provided back to Inland Revenue for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support Inland Revenue's core operational requirements. Careful consideration has been given to the privacy, security and confidentiality issues associated with using tax data in this project. In particular, in the IBULDD dataset, individuals' tax data has been aggregated to the firm-level. Furthermore, only people authorised by the Statistics Act 1975 are allowed to see data about a particular firm.

Abstract

This paper investigates the microeconomic dynamics of the New Zealand economy using a powerful new SNZ-held dataset. For the first time, tax data covering operating performance and position (IR10), company income declarations (IR4) and sales & purchases (GST) have been combined with firm-level (LEED) employment data and Customs merchandise trade data to create a full coverage Longitudinal Business Database (LBD). We use this data to expand the available set of firm performance measures to include multi-factor productivity, profitability and export intensity. These and other performance variables are used to examine the distribution and dynamic of New Zealand firm performance, focussing on the characteristics of firms that display superior or inferior firm-level outcomes.

JEL Classifications D21; O12

Keywords: firm dynamics; productivity; profitability; exporting; foreign ownership

Executive Summary

The New Zealand economy is a complex system whose operation cannot be fully understood by pondering macroeconomic statistics. The policy process can be greatly improved by developing a deeper understanding of the microeconomic dynamics of the economy– an understanding that is partly delivered by research using detailed unit-record data. Statistics New Zealand (SNZ), together with the Inland Revenue Department (IRD), has been instrumental in enabling work of this sort to be done in New Zealand. Most recently, this collaboration has produced a prototype Longitudinal Business Database (LBD), integrating a plethora of survey, tax and Customs trade data within the SNZ secure environment. From a research perspective, the breadth of data included within the prototype LBD enables advances to be made in many of the microeconomic studies previously investigated in New Zealand, as well as opening up many new avenues for investigation. The purpose of this paper is to explore and highlight some of this potential.

At this stage, the analysis presented provides motivation for more work, not definitive answers to questions. However, some clear themes arise from the data:

- Strong “Darwinian” processes act on firms, weeding out the weak and rewarding the strong. Having said that, there is great variation in firm performance within industries, implying that economic models assuming homogeneous producers, or rhetoric labelling particular industries as “good” and others “bad”, may be somewhat counterproductive. Low (high) average productivity industries always contain high (low) performing firms; and
- Firms with international connections (exporting or foreign-ownership) have a clear productivity advantage over purely domestic firms. Initial exploratory work suggests that the performance advantage exporters have exists prior to their entering exporting (ie, firms self-select into exporting).

The dataset stands out – internationally – for both its comprehensive coverage of firms and the sheer variety of data captured. The unique strengths of this data bode well for the ongoing research programme, which is focussed on topics close to the heart of the public policy debate on the state of the economy.

Table of Contents

List of Tables

List of Figures

Some rise by sin, and some by virtue fall:[1] Firm dynamics, market structure and performance

1. Introduction

The New Zealand economy is a complex system whose operation cannot be fully understood by pondering macroeconomic statistics. With recent advances in data availability, policymakers have benefited from a deeper microeconomic understanding of the dynamics of the New Zealand economy. Much of the work done in this area has focused on the key economic growth parameters of labour productivity (value-added per unit of labour) and employment. Examples include understanding the contribution of firm dynamics to employment growth (Carroll et al. 2002), productivity growth (Law & McLellan 2005), the distribution of firm size (Dixon et al. 2005), the effects of agglomeration (Maré & Timmins 2006), and the role of employer-employee matching (Maré & Hyslop 2006). Analyses of this sort are becoming common throughout the OECD, helping to benchmark market dynamics across countries and shape international understanding of the appropriate policy settings for economic growth (eg, Ahn 2001, Scarpetta et al. 2002, OECD 2004).

Statistics New Zealand (SNZ), together with the Inland Revenue Department (IRD), has been instrumental in enabling work of this sort to be done in New Zealand by allowing the relevant administrative & survey data to be accessed in a way that protects the privacy and confidentiality of individuals and businesses. A key focus of

[1] From *Measure for measure* by William Shakespeare.

SNZ's work has been to determine whether they can provide timely new official statistics on the microeconomic performance of New Zealand businesses using a Longitudinal Business Database (LBD).[2]

From a research perspective, the breadth of data included within the prototype LBD enables advances to be made in many of the microeconomic studies previously investigated in New Zealand, as well as opening up many new avenues for investigation. The purpose of this paper is to explore and highlight some of this potential. Section 2 outlines the contents of the LBD, while section 3 discusses the choice of performance metrics used in this paper. Section 4 presents a selection of outputs from the data, with reasons for being cautious about those outputs left to section 5. Section 6 concludes by reiterating the strengths of the data and briefly outlining potential future work.

2. Description of the dataset

SNZ's proposed functions for the LBD necessitate reasonably full coverage data. In general, such data is either held on SNZ's Business Frame (BF) or derived from administrative data held by other government departments. The core administrative data on the LBD currently consists of the Longitudinal Business Frame (LBF) with goods & services tax (GST) returns, financial accounts (IR10), and company income tax returns (IR4) provided by IRD; information on employers and employees aggregated to the firm level (sourced from IRD via LEED[3]); and shipment-level merchandise export and import data provided by Customs. The nature of each of these datasets is briefly discussed below.

As its name suggests, the LBF is a by-product of SNZ's sampling frame (the BF) and contains longitudinal information (eg, industry, ownership type, and sector) on a wide population of firms.[4] The quality of the LBF's representation of firm characteristics, and changes in those characteristics, is a function of the maintenance processes for the BF, the ability of respondents to answer survey questions, and the quality of

[2] This dataset has been constructed under the working title of IBULDD (the acronym for SNZ's project to construct the LBD).
[3] The Linked Employer-Employee Dataset.
[4] The business frame is a set of history tables that record changes in firm characteristics, whereas the longitudinal business frame is a monthly "unwinding" of those history tables to reflect the actual timeseries characteristics of the firms on the business frame.

supplementary sources used. GST data is used to help maintain the accuracy of the BF (particularly to track the births and deaths of firms) and, consequently, a significance threshold exists at the mandatory GST filing level, below which BF coverage is limited. Large economic units are surveyed either annually or triennially to maintain the accuracy of the data held. The LBD version of the LBF holds data from April 1999 to June 2007.

GST data is collected on a monthly, bi-monthly or six-monthly basis by IRD, depending on the size of the firm filing. GST data include information on sales & purchases. SNZ manipulate this raw data to create the Business Activity Indicator (BAI) dataset (also included in the LBD). The primary manipulations applied to generate the BAI data are to temporally apportion the GST data down to a monthly frequency, apportion returns across GST group members, and apply limited imputation in cases where a single return appears to be missing. In the LBD BAI data is available from April 1992 to May 2007.

IR10 data is essentially a set of company accounts composed of a statement of financial performance and financial position. Consequently this form contains information on sales (and other income) and purchases, as well as a detailed breakdown of expenditure including depreciation, research and development, and salaries & wages. Balance sheet items include the usual suspects: fixed assets (broken down into vehicles; plant & machinery; furniture & fittings; land & buildings; and other), liabilities broken down into current & term, and shareholders funds. IR10s are available for the 1998/99 to 2005/06 financial years.

Like IR10s, IR4 returns are available on the LBD for 1998/99 to 2005/06 financial years. IR4s are declarations of taxable income for companies and, as such, include variables on overseas income, interest & dividends & income from “business or rental activities”. They also contain a binary foreign-ownership indicator.[5]

LEED data is constructed by SNZ from IRD tax data, notably Pay-As-You-Earn (PAYE) returns for employees. To protect the confidentiality of individuals, LEED

[5] A foreign ownership indicator (percentage of the firm owned offshore) is also held on the BF (LBF). The advantage of the IR4 indicator is its greatly superior coverage for companies and potentially more timely updating.

variables available in the LBD dataset have been aggregated to the firm-level (allowing the data to be accessed through the Datalab). Variables available in this manner include counts of employers (on an annual basis) and employees (on a monthly basis) with matching data on income. Summary characteristics of individuals also include gender and banded age breakdowns, tenure distributions of employees, and summary measures of the dispersion of wages within the firm. Accessions and separations are summarised at the firm level, as are counts of contractors employed (with remuneration).

Customs data is linked to the LBF initially via probabilistic matching with subsequent manual matching for any remaining unmatched large-value Customs clients (Smart & Johnstone 2007).[6] The dataset contains daily shipment-level information from January 1988 through to October 2007 covering goods (defined by the 10-digit harmonised system, HS10), countries of origin and destination, values, volumes, weights, currency of trade, port of entry/exit and mode of transportation.[7]

In addition, a number of SNZ sample surveys have been linked to the LBD, namely:

- Annual Enterprise Survey (AES) 1997-2006;
- R&D Survey biennially 1996-2006;
- Business Practices Survey (BPS) 2001;
- Innovation Survey 2003;
- Business Finance Survey (BFS) 2004; and
- Business Operations Survey (BOS) 2005-2006.

Being sample surveys, these data are relatively sparse in the LBD. Other than AES, these datasets are not used in the current paper, and interested readers can find detailed descriptions of the survey collections on SNZ's website. AES is SNZ's primary data source for the production of National Accounts, and as such is the benchmark dataset for estimation of value-added. The survey is full coverage for large firms with a stratified sample survey for smaller firms, and has industry-specific questions in order to accurately measure aggregate gross domestic product. In this

[6] This process results in over 99% of the value of free-on-board (fob) merchandise exports being matched to the LBF for each year between 1999/00-05/06. The quality of the match between Customs clients and firms deteriorates prior to the mid-1990s.
[7] Some variables are not available in earlier years because of changes to data capture processes.

paper we use AES postal responses to assess the accuracy of our value-added measure derived from tax sources.

Lists of firms that have received assistance from government agencies, together with information on the size and nature of the assistance, have also been probabilistically matched (on contact details) to the dataset to enable evaluation of these schemes.[8]

Some choices have to be made about the relevant population for the statistics produced in this paper. First, we choose our unit of observation as the enterprise (referred to as the firm throughout this paper). Much research in this area uses the plant (or geographic unit in SNZ's nomenclature) as the unit of observation. However, in New Zealand data most financial variables are only observed at the firm (or tax reporting) level, not at the individual plant (the main exception being LEED salaries & wages). To avoid the issues inherent in apportioning output to firms with multiple locations, this paper focuses on firm-level performance metrics. From a conceptual perspective the span-of-control covered by a firm may be more appropriate to the types of analysis expected of the LBD. For example, business performance surveys (such as BOS) are generally targeted at the firm using the logic that firm practices are expected to be set at this level of organisation.

Second, the time frame of longitudinal analysis involving all data sources is limited by the availability of LEED data. At the time the results in this paper were prepared, full data was only available for the six years financial years from 1999/00-2004/05. An annual frequency is imposed on the data by the IR10, IR4 and working proprietor tax returns. All sub-annual data (Customs, BAI, LEED employee data) is annualised to each firm's financial year and then allocated to the "notional" 31st March year-end that has the greatest overlap with the financial year.[9]

Third, we have to define an in-scope firm. To simplify the discussion of data coverage and to increase the likely applicability of the performance metrics estimated, we

[8] Agencies supplying data are the Foundation for Research, Science & Technology; New Zealand Trade & Enterprise; the Ministry of Tourism; Te Puni Kōkiri; and the Ministry of Social Development.
[9] In practice, due to IRD requirements, most firms actually have a 31st March balance date.

include only “private-for-profit” firms,[10] and additionally exclude households, ANZSIC Division M (Government Administration & Defence) and firms not located in New Zealand. For practical reasons, “firms” that have never reached the BF materiality threshold and, therefore, do not appear in the LBF are excluded from the analysis (as they are not currently assigned to industries). Similarly a small number of firms that are on the LBF, but have partial or no ANZSIC information, are dropped from the analysis.

Finally, we must determine criteria for whether we treat a firm as active in any particular year. SNZ’s standard approach is to define populations using the dual criteria of “live” and “economically significant”. The latter criteria relates to materiality, while the former assesses whether the business is in operation. Variables capturing these criteria are located on the BF (and LBF) which, in turn, makes use of IRD data to maintain the accuracy of the population characteristics. However, through the LBD we have access to a wider set of administrative data from which to assess business activity. Naturally, the use of this wider set of data increases the potential to observe active businesses. We define an “economically active” (ie, in-scope) firm as one where we observe output, purchases of inputs or factors of production, specifically: positive employee count or PAYE salaries & wages; positive BAI sales or purchases; and/or positive IR10 total income, total expenditure or total fixed assets. This sets the population much wider than a live & economically significant approach, primarily because the economically active rule does not have an explicit materiality threshold,[11] and because the additional tax data suggests some firms be treated as active despite being ceased on the BF.

Table 1 sets out the size of our population in each year, together with entry and exit rates defined by a firm being active in one year, but not in the relevant adjacent year. Even in this simple breakdown, there is much dynamism present with approximately a fifth of the population of firms either entering or exiting in a given year. Put another way, there are 687,573 distinct firms within the dataset with roughly two thirds of

[10] Defined loosely as business types 1-6: individual proprietorship; partnership; limited liability company; co-operative company; joint venture & consortia; and branches of companies incorporated overseas.
[11] The mandatory filing threshold for GST provides an implicit materiality threshold, since firms that do not reach this threshold may not file and therefore have the activity observed; or if they do file, they may not be coded to an industry on the BF (and are, therefore, excluded from the population).

them active in any single year. Table 2 sets out the patterns of activity present in the data. A small proportion of the observed firm turnover is due to firms that enter and exit the population on an intermittent basis, and it might be reasonable to expect that some of these transitions are spurious.[12] However, 95.9% of firms experience a single continuous spell of economic activity,[13] with 39.0% of firms in the dataset continuously economically active over the full period. Overall, the general picture of firm dynamics is consistent with survival analyses previously published using more "traditional" population definitions (eg, MED, SNZ various years).

Having set the population characteristics, it is necessary to discuss missing data. In this paper, we assume that missing employment (working proprietor) data implies zero employees (working proprietors) on the grounds that personal income tax non-compliance is likely to be negligible in the population of firms that meet the mandatory GST filing threshold. Similarly it is assumed that Customs data is comprehensive.[14] For this exploratory analysis, we do not make any attempt to impute missing data in other datasets. Tables 3 & 4 set out coverage rates for each of our administrative datasets by firm size & industry respectively. Administrative data can be missing for a number of reasons, including:

- Filing is not mandatory. In terms of the potential for bias to be introduced into the analysis, two issues stand out from tables 3 & 4: For BAI, missing data largely arises because of GST exempt financial activities in the finance & insurance industry; and IR4s are company returns and therefore not filed by other business types, explaining very low reporting rates in some industries;[15]
- Filing is mandatory, but a firm is non-compliant (non-compliance with GST reporting appears very low);
- Data is filed, but has to be discarded because it is of insufficient quality for statistical purposes. In the case of IR10s, a large number of missing

[12] As a consequence table 1 is not particularly robust to variations in our definition of entry. For example if our definition of entrants were to exclude firms that exit the immediately following year (on the grounds that some may be spurious "firms") then our entry rate would drop by approximately 9%.
[13] Ignoring issues of left and right censoring.
[14] Bearing in mind the fact that Customs exports less than $1000 may not be captured, and that probabilistic matching will naturally yield some small level of false negatives.
[15] The LBD holds IR4 returns for 79% of economically active companies, where companies are defined by the BF business type. In a very small number of cases the BF business type is inconsistent with the fact that the firm has filed an IR4 (eg, less than 0.1% of IR4 filers have a recorded business type of sole proprietor or partnership). In such cases, we assume that the filing of the IR4 implies that the firm is indeed a company (ie, we ignore the contrary evidence on the BF).

observations exist because a returned form only contains zeros or fails simple internal consistency checks (eg, that totals "approximately" sum correctly);[16]

- One data source incorrectly implies a firm is economically active, thus giving the impression that other data should be present. For example, there is undercoverage of both BAI and IR10 data for entering and exiting firms, which may be reflective of incorrectly assessing the timing of entry and exit; or
- Links between IRD & BF firm identifiers are missed, partial or incorrectly apportioned across the enterprises that the filing covers. The rate at which this occurs is assumed to be low.

3. Performance measurement

This paper focuses on a small number of performance variables, namely sales, total employment, merchandise exporting, profitability, labour productivity and multi-factor productivity (MFP).[17] The first three are trivially calculated, respectively, as BAI total sales; an average of the twelve monthly (PAYE) employee counts in the year[18] combined with a count of working proprietors from LEED; and free-on-board Customs exports. Profitability is measured as the ratio of (IR10) taxable profit to (BAI) sales. Our productivity measures require the construction of value-added, defined as gross output less intermediate consumption, and approximated by:

$$VA = \text{sales} - (\text{purchases} - \Delta\text{stocks}) \qquad (1)$$

where sales and purchases are sourced from the BAI & changes in stocks are sourced from IR10s.[19,20] BAI data is used for sales and purchase data due to concerns over under-reporting of IR10 purchases (Cox 2006), and because BAI

[16] Because IR10 financial performance and position data can be assessed as passing these "edit checks" independently, it is possible that only one of the two is missing. In this paper we report only rates of coverage for financial performance data because we do not make use of any asset or liability data in this paper (for the record, financial position data is more likely to fail edit checks).
[17] Because IR10 data constitutes a reasonably full set of financial accounts, the data would allow the construction of additional financial performance measures such as financial solvency. These are not covered in the paper for the sake of brevity, but should be considered as potential avenues for further research.
[18] Known as rolling mean employment (RME). The measure of RME used in this paper differs from that used in official LEED outputs as it excludes working proprietors that receive PAYE income (to avoid double counting of their employment).
[19] GST is removed from sales & purchases data so that value-added is estimated at factor cost, rather than market prices. Most other financial data, including survey responses, are collected on a GST-exclusive basis.
[20] All results in this paper are nominal. Potential methods for adjusting to real output all involve applying industry average input and output producer price indices that may help conceal data issues. Appropriate adjustment methods are discussed in the caveat section.

coverage is superior.[21] Prior micro analysis has not had access to IR10 data, and thus the stock adjustment has not factored into earlier firm-level productivity calculations. The effect of this adjustment is, in general, minor. For approximately three fifths (61.5%) of value-added observations the stock adjustment is zero, almost exclusively because opening and closing stocks are both reported as zero. Weighted by total employment, the mean (median) relative contribution of the stock adjustment is 7.2% (0.2%) of value-added.[22] The correlation between the labour productivity measures with and without a stock adjustment is 0.959 in levels and 0.903 in growth rates. All this suggests that, while improving the conceptual accuracy of the value-added measure, such an adjustment is unlikely to undermine the results of previous authors who have been unable to make such an adjustment. We retain the adjustment in the remainder of the paper, noting that the use of IR10 data decreases the number of observations of value-added.

Table 5 sets out the number of productivity observations we have. Initially we lose 28.5%[23] of observations simply from the fact that many firms have zero employment – that is, they have neither employees nor working proprietors. A large number of these zero employment firms are in the finance & insurance, and property & business services industries (table 4), perhaps a sign of a large number of "shell" or asset-holding companies in our data. Next we lose a relatively modest 3.0% of observations from the absence of BAI data. As we noted in the prior paragraph, another major loss of observations (15.0%) comes from requiring the stock adjustment to the labour productivity calculation. Finally, because distributions of firm performance are highly skewed, labour productivity is reported as the log difference between value-added and employment. Taking logs of value-added results in another 9.6% of observations being dropped from labour productivity calculations because value-added is zero or negative. As with missing data, non-positive value-added is disproportionately associated with entering and exiting firms. ANZSIC divisions A, B,

[21] Comparison of similar variables across data sources often yields some inconsistency. It is hard, in such cases, to disentangle the relative effects of definitional & timing differences, the respondent's ability to answer accurately, and other sources of potential error such as apportionment or aggregation. Work within the IBULDD project suggests that most IRD-sourced variables correlate well to like variables collected through postal AES returns (Smith 2006).

[22] Calculated as $|\Delta stock|/(|\Delta stock|+|sales-purchases|)$. Unweighted, the mean (median) contribution is 8.3% (0.0%). Not surprisingly, the importance of the stock adjustment is concentrated in wholesale & retail trade; agriculture, forestry and fishing; and manufacturing.

[23] These percentages are susceptible to the order in which they are calculated.

D & K also have higher rates of negative value-added.[24] Overall, we are left with 1,228,322 observations of labour productivity (corresponding to 43.9% of economically active firms).

Table 6 shows correlations comparing our key value-added measure against the measure derived from postal responses to AES.[25] In general the correlation of log-levels is very respectable with the finance and insurance industry showing the weakest correlation (at 0.6227). Turning to growth rates, we find that both short-term and longer-term growth rates are more weakly linked across data sources.[26] Growth rate comparison is made difficult by the selective nature of any AES longitudinal sample (biased towards the largest firms). However, for industries where large numbers of observations are available the four-year growth rates show significant positive correlation across the data sources. Overall, the results in table 6 give us some confidence that the LBD value-added variable is plausible and fit for research purposes.

MFP is calculated by way of regression assuming a Cobb-Douglas production function in labour (RME) and depreciation expenses (from IR10s) with industry-specific coefficients (a mix of one- & two-digit ANZSIC), year-specific dummies, and the potential for non-constant returns to scale. MFP is the residual of this estimation with industry average and year effects added back. That is, MFP is the component of value-added that is not explained in our model by capital and labour inputs.

The use of depreciation costs rather than a "true" capital services measure is forced on us by the absence of capital stock (and/or capital investment) data.[27] The number of MFP observations is lower (38.8%) due to reported zero depreciation (table 5). We

[24] ANZSIC A is Agriculture, Forestry & Fishing; ANZSIC B is Mining; D is Construction; ANZSIC K is Finance & Insurance.
[25] Due to sample size, all years are pooled together. Because the comparison is on log levels and differences we lose 14,997 observations with non-positive value-added in at least one of the measures (in 67% of dropped cases the AES measure is non-positive, similarly 52% for the LBD measure).
[26] To attempt to smooth potential short-term mismeasurement issues, the four year growth rate is calculated as the difference between the two-year average of value-added in the last and first years of the dataset.
[27] We do have closing book values of fixed assets. To a certain extent, the use of depreciation costs should not be too problematic since tax-deductible depreciation rates in New Zealand tend to have an economic basis. However, our depreciation cost should be supplemented by the cost of debt or equity to derive a full cost of capital. We assume this full cost of capital to be correlated to the depreciation cost and, therefore, ignorable in this preliminary paper. A cost of capital component will be investigated in future.

lose 7.8% of the value-added and 7.5% of the employment associated with the labour productivity measure (ie, of the sub-population with positive value-added).

Appendix A presents the regression coefficients from the MFP calculation. Some coefficients seem implausible, particularly because of the relative contributions of labour and capital, and the implied rate of increasing returns in some industries. Specifically, we might expect the contribution of capital would be higher (a third being a ballpark figure from macroeconomic estimates) with only mildly increasing returns to scale, together implying lower labour coefficients than those estimated. For this exploratory paper, we treat these estimates as adequate, noting that the MFP calculation needs further investigation.[28] One approach to be looked at in more detail is using alternative specifications, particularly a generalised CES production function (see, for example, Grimes 1983).

4. Analysis

We begin our analysis by looking at the correlations across performance metrics. Table 7 shows (Pearson & Spearman rank) correlations across our three core levels measures of profitability, labour productivity & MFP. As we would expect, the rank correlations are positive & significantly different from zero.[29] Presumably the high correlation between labour productivity & MFP is partly a reflection of any inadequacy in our Cobb-Douglas model and/or the high degree of correlation between capital and labour. Table 8 shows correlations of annual growth rates, as opposed to levels and expands the performance measures to include sales & employment.[30] Again, most measures are positively correlated, with the main exception being the relationship between employment growth and productivity growth. This negative correlation is perhaps to be expected over the short-run, consistent with an economic model with adjustment costs (ie, as firms scale up this requires changes in structure, learning, etc, which impose costs). Such a model is also consistent with our finding that sales and employment growth are positively correlated (ie, a scale effect).

[28] In a parallel analysis, Dixon (2007) found similar results using an unbalanced panel of AES/IR10 data with total fixed assets as the capital input and aggregate, rather than industry-specific, coefficients.
[29] It is perhaps better to compare Spearman rank (as opposed to Pearson) correlations for the profitability measure given its greater susceptibility to generating extreme values in the distribution.
[30] Throughout this paper, growth rates are measured as log differences for all variables except profitability, where a simple difference is used.

Figures 1, 2 & 3 present industry-level distributions of labour productivity, MFP and profitability (respectively) in 2005.[31] Looking at the productivity distributions first, it appears that some industry differences are partially explained by differing average capital intensities. For example, the mining industry (ANZSIC B) productivity distribution sits to the right in Figure 1. Conversely the accommodation, cafés & restaurants industry (ANZSIC H) sits to the left of the labour productivity distribution. After controlling for capital intensity the apparent productivity differences between these two industries is diminished (Figure 2). Most industry profitability distributions are centred close to zero (Figure 3), with communication services and health and community services being the most noticeably right-skewed.[32] Conversely, decent returns in 2005 seemed hardest to come by in the accommodation, cafés & restaurants industry.[33]

Despite the high degree of churn in the underlying population of firms suggested by Table 1, distributions of performance levels and growth rates are remarkably stable over time. For example, Figure 4 shows the distribution of profitability for three years (2001, 2003 & 2005). The distribution across years is very similar for negative profits, with moderate year-on-year variations in the proportions of firms that record positive profitability. Similarly Figure 5 shows that annual growth rate distributions are also quite stable across years. Having said that, the range of annual growth outcomes is broad with log differences in labour productivity across years often exceeding one.

Attempting to reconcile this vibrant firm-level dynamism with the seemingly structural stability in aggregate distributions has a long history in the economics literature (see Sutton 1997 for an excellent review). Part of the reconciliation has to do with the persistence of performance in incumbent firms, remembering that these firms account for almost two fifths of observations. Table 9 summarises the transitions of firms between 2000 and 2005 labour productivity deciles.[34] The main point to note is

[31] Productivity distributions have had one percent of the density at each end excluded to minimise risks around outlier disclosure, and to focus attention on the bulk of the distribution. Profitability distributions have had 5% removed from each end.
[32] It is perhaps worth reiterating that these graphs are unweighted distributions of firms, so the correct interpretation of this Figure is that a significant proportion of firms in these industries have higher profitability.
[33] Looking at Figure 3, several industries appear to vie for the title of least profitable. This industry is singled out by considering mean profitability levels (results not reported).
[34] Decile boundaries in each period are determined with reference to all active firms in the respective period as opposed to the subpopulation that appear in both periods, which explains why there are not an equal number of firms in each decile.

that a large proportion of surviving firms maintain their relative productivity level. Overall, 23% stay in same decile while a further 30% move only one decile up or down the distribution. Another way to think about this persistence of performance is to look at the autocorrelation of various performance metrics across various lag lengths. Table 10 demonstrates the strength of the temporal relationship between the levels of our three key performance metrics, labour productivity, MFP & profitability. Looking at the autocorrelations in annual growth rates it appears that there is some short-term reversion in performance (ie, "good" years are followed immediately by "bad" years). Over the longer term, annual growth rates do not appear to be correlated at all (except in the case of employment growth, where a weak negative correlation persists). These results concur with work done previously (see, eg, Law & McLellan 2005 & Law et al. 2006).

This is not to undermine the important contribution to the economy that comes from entering and exiting firms. As Table 11 and previous analyses (eg, Carroll et al. 2002) have demonstrated, entering and exiting firms account for a large proportion of net job creation. Using the LBD we can describe the productivity impact of firm turnover. Standard methods of decomposing the contribution of firm turnover to productivity usually consider impacts over a five year period. Since this would leave us with a single observation, we instead, in Figures 6 & 7, present cohort analyses for entering and exiting firms from which it may be inferred that entry and exit probably make the expected long-term positive contribution to aggregate productivity growth. Entrants have lower productivity initially and then tend to move to or past the average productivity level of full-period incumbents by their second or third year of operation (conditional on survival).[35] Conversely, exiting firms have below average productivity throughout the last few years of their existence.

Two further groups of firms have a particular interest for innovation, trade & competition policy: exporters & foreign-owned firms. We turn now to a brief discussion of each of these groups. Very little is known about the microdynamics of New Zealand's export sector. Because our measure of exporting comes purely from

[35] These cohorts are restricted to firms that appear to have a single continuous period of economic activity (ie, intermittent firms are excluded). This restriction does not affect the bulk properties of the results, but simplifies the interpretation of cohort performance.

Customs merchandise trade data, this section of the paper focuses exclusively on manufacturing firms. We do this to reduce the potential of misestimating the correct denominator in our calculation of the proportion of firms that export (an issue that is sometimes overlooked when very low rates of exporting are reported in New Zealand), and to control for broad industry in our discussion of relative productivity performance.

Previous research suggests that exporting is concentrated in a small number of firms, and that an even smaller number of firms generate a large proportion of sales from exports. For example, Simmons (2002) reports the proportion of all firms that export between 4-5% (a figure susceptible to the "appropriate denominator" criticism). In our data we find that 11.4% of manufacturers exported goods in 2005. Figure 8 summarises export intensity deciles in 2005 after the 88.6% of firms that have zero exports have been removed.[36] The data strongly supports the idea that exporting is a sideline activity for most firms, with the mean (median) firm exports constituting 17.9% (5.9%) of total sales.

Figure 9 demonstrates that, while relatively scarce, manufacturing exporters punch above their weight displaying higher average labour productivity. Simple tests of differences in means (1% significance level) suggest that both incumbent & new manufacturing exporters have higher labour productivity levels than non-exporters (and incumbent exporters are significantly higher than entrants also), and that entering exporters have higher annual employment growth than non-exporters (starting from a higher average total employment). The theoretical (and some empirical) literature suggests that exporters may experience faster productivity growth through, for example, learning effects. These effects might be expected to be more prominent in market entrants, rather than incumbent exporters. To investigate this possibility, Figure 10 breaks the labour productivity growth distributions into entering exporters, incumbent exporters, and non-exporters. Labour productivity growth in exporters (neither incumbent nor entering) is no higher than in non-

[36] A further 157 firms are excluded because their total exports appear to exceed their total sales. Three potential reasons that Customs export values could exceed BAI sales are: timing differences in reporting; false positives in the probabilistic matching of Customs records; and/or apportionment of BAI sales within GST groups. It could be that, on further investigation of the data, these firms increase the rank of the more intensive exporters.

exporters.[37] Potentially, we should be looking for longer-run effects from foreign market participation. However, these preliminary findings are consistent with the international literature, for example, Bernard & Jensen (1999) who find that employment growth is higher in US exporters, but not labour productivity growth.

Figures 11 & 12 compare the labour productivity distributions of foreign-owned and domestic firms using the IR4 foreign-ownership indicator as the basis of splitting the sample. This analysis is restricted to limited liability companies, since only these firms file IR4s. Focussing first on Figure 11, it is apparent from the data that foreign-owned companies are more productive than domestically-owned equivalents, with the difference in productivity levels quite startling. Perhaps this is a consequence of the simple univariate breakdown in the data? In particular, the foreign-owned firms are concentrated mainly in five industries that also tend to have higher mean labour productivity levels: mining; construction; wholesale trade; communication services; and finance & insurance. Another possibility is that foreign-owned firms are also exporters. Foreign-owned manufacturers are roughly four times more likely to be goods exporters than domestically-owned manufacturers.

Figure 12 breaks the labour productivity distribution of manufacturing companies down by both ownership and export status. Apparently, exporting behaviour cannot explain the size of the productivity gap between foreign-owned & domestic-owned manufacturers. All this suggests that careful econometric analysis is required to disentangle the underlying causes of higher firm productivity. For example Figure 12 is consistent with a model where FDI has no impact on productivity, but rather that foreign firms have better access to capital and can buy out more productive NZ firms; or, alternatively, that the effect of FDI is on organisational behaviour rather than access to foreign markets. Before turning to a discussion of future work (in our concluding section), we quickly outline data issues that also need to be factored into our interpretation of these results.

[37] The resulting picture is somewhat different if BAI zero-rated sales is used as the measure of exporting (as we shall see in the next section) emphasising the importance of understanding the origins of the data.

5. Key caveats on the data

The prior discussion ignores many issues of data quality. The largest of these issues are summarised in this section. While this section of the paper appears somewhat daunting, none of the caveats raised present insurmountable obstacles to the use of the data provided appropriate cautions are attached to outputs. Where feasible, next steps are discussed for researchers wishing to remove these obstacles.

5.1 Longitudinal firm continuity

We would like the LBD firm to correspond to an economic definition of a firm, such as *the combination of production factors within the span of control of a set of owners*. In a cross-sectional sense the Business Frame enterprise satisfies this definition nicely. However, longitudinally, this relationship tends to be weaker because SNZ tracks the continuation of legal units, not firms. Eurostat sets out three criteria for measuring firm continuity – control, economic activity & location – and require two of these to remain the same for a firm to be described as continuing (Eurostat 2003). However, new legal units may be created on the BF without any of these three continuance criteria being violated. There is scope within the future development of the LBD for some repairs to be executed on the longitudinal continuity of firms. This work would improve outputs looking at firm entry and exit, possibly yielding insights into differences between greenfields and mergers & acquisition start-ups (see, eg, Baldwin & Gu 2006).

On a related issue, some caution is necessary around the interpretation of entering and exiting firm productivity due to the inability of the data to accurately discern exact dates that production starts and stops. For example, in Figure 6 and 7, the sharp pick-up (drop-off) in productivity of entrants (exiters) in their first (last) year is perhaps indicative of data issues that could influence the productivity calculation, for example, divergence between timing of administrative tasks such as GST registration and the production of goods & services and, similarly, winding down of financial accounts at business exit. As suggested earlier, these issues are a likely cause of higher rates of missing tax data in years that firms enter and exit the population. Our approach of an inclusive economic activity-based population is also likely to exacerbate these issues by including firms that have very low output.

5.2 Employment measurement

The estimation of entering and exiting firm labour productivity is further exacerbated by the fact that working proprietor data is most often only observed annually (ie, there are no part-year counts). The mean (median) entering firm has an RME of 1.64 (0) and working proprietor count of 0.95 (1). Thus the current assumption that working proprietors work the full year has a measurable impact on labour productivity estimates in the first year of activity (and, similarly, the last year). Put another way, if we assumed that working proprietors only worked half the year of start-up and/or exit, the estimated mean labour productivity of entrants (exiters) would exceed the incumbent labour productivity in the year of entry (exit). On the positive side, the inclusion of RME as the labour input accounts for mid-year start-ups/shutdowns, compared to prior studies, which have relied on BF annual snap-shot employment and thus have to assume a labour input pattern over the year for both employees and employers (eg, Maré & Timmins 2006, who carefully test their estimates using both the "full-year" and "half-year" assumptions).[38]

The fact that the employment data involves simple headcounts will also have a tendency to overestimate labour input (because of part-time workers[39]). In the absence of detailed hours worked data, the most common approach to correcting for this issue has been to adjust counts by industry-level average hours worked sourced from either SNZ's Household Labour Force Survey or Quarterly Employment Survey. Such adjustment only improves the comparison across, rather than within, industries, and can usually only be done with confidence at the two- or three-digit industry level.

In addition, some thought should be given to whether at least some part-time employment proxy could be established in the data (see Maré & Hyslop 2006 for an example of how this has been done).[40] A further issue arises for working proprietors in that some owners of firms will receive taxable income purely as a return on equity,

[38] Potentially the monthly employee count could be used as a basis for modelling the period over which working proprietors are actually working, though for 61% of entering firms with non-zero total employment, RME is zero.
[39] Though the RME component does account for part-year workers.
[40] It may also be desirable, for some applications, to attempt to adjust for labour quality as well as quantity. Progress could be made through the use of the available aggregated employee characteristics (age, gender & tenure). See, for example, SNZ's recent investigation on this for official productivity statistics (McNaughton 2006). As an alternative, salaries & wages could be used as the labour input measure. On the positive side, S&W could reasonably be assumed to incorporate both hours & quality adjustment factors. On the negative side, S&W may incorporate elements that could equally be attributed to firm value-added (eg, employee-firm matching effects).

without any labour input being supplied at all. Identifying this subset of owners is difficult.[41]

5.3 Deflators

As noted earlier, all results in this paper are presented in nominal terms. It is usual, in policy applications, to be primarily interested in real productivity growth. However, as with measuring hours worked, no input or output prices exist at a comprehensive firm level and it is usual to apply industry-level input & output producer price indices to improve the cross-industry comparison of firm performance.

5.4 Capital data in BAI sales and purchases

The GST-based sales (and purchases) data is potentially contaminated by capital income (expenditure). As the BAI documentation notes:

> "…GST sales variable includes other items such as: Sales of second-hand assets… [and] sales of businesses themselves. If they are sold as a going concern the sale is zero-rated. The amount of the sale will still appear in the GST sales variable.
>
> …GST purchases variable also includes: Purchases of land, buildings, plant and machinery etc... [and] purchases of businesses themselves. If the business is sold as a going concern the amount of the sale is not record[ed] as a GST purchase." SNZ (2001)

In a particular year, this capital data could potentially swamp measurement of true firm value-added (productivity). BAI processes to address this issue are only targeted at removing large spikes in values that might affect firm confidentiality in reported outputs.

One area where this capital contamination appears to manifest is in the use of zero-rated sales as a proxy for export earnings (since exports do not attract GST). An earlier version of this paper (Fabling 2007a) used positive zero-rated sales as a proxy indicator of exporting behaviour and found implausibly high growth rates in labour productivity for exporting firms (relative to non-exporters). The figure in

[41] There is potential to use comparison with BOS full & part-time working proprietor data to estimate the size of the problem. However, the population of BOS has a minimum size cut-off of six RME. It is not clear whether small employment firms are more or less likely to have "sleeping" partners.

Appendix B replicates that earlier analysis for manufacturers in our population. As Fabling concluded:

> "It appears from this data that entering exporters drive the difference in growth rates. However, a perhaps more realistic explanation of the high growth rates is that the [BAI-based] export indicator is a poor proxy for measuring true exporting behaviour…"[42]

Obviously capital contamination is most influential in analyses where the population is restricted to firms with zero-rated GST sales. Fortunately, the rate of entry into (exit out of) zero-rated sales is quite low, with some of that activity presumably related to true export activity. Overall the number of productivity observations that are affected may be small, though this is hard to estimate with certainty without a comprehensive exports measure (ie, including service exports). There is potential to further investigate the importance of the capital contamination issue using IR10 data on sales, gains/losses on disposals of fixed assets and book values of fixed assets, together with better export identification as discussed below (and, perhaps, data held on the LBF regarding transfers of plants between firms).

5.5 Exporter identification

Where the BAI does have an advantage over Customs data is that it should capture trade in services – clearly an important subject for analysis. One potential way to identify service exporters (that does not rely on BAI data) might be to use IR4 foreign income data combined with BF trade in services (balance of payments) indicators and Customs data.

Though we have not discussed it in detail in this paper, we have a number of exporters in our dataset that, according to our ANZSIC classification should not be in the business of substantially exporting goods, but are. These firms may be "head offices" (explaining their service-related ANZSIC) in enterprise groups that contain a firm that should, more appropriately, be associated with the exported good (eg, a manufacturing subsidiary). The other way in which group structures create problems is when groups restructure and the exports appear to shift between firms in the group. This issue has direct parallels with the discussion of false entry and exit above.

[42] We could also add to the apparent evidence for capital contamination that 45% of these apparent entering exporters are not exporting in the following year, though as Fabling & Sanderson (2008) indicate, this sort of intermittent exporting may not be atypical behaviour.

5.6 Missing data

Section 2 of the paper noted a number of potential sources of bias arising from missing administrative data. As Table 5 demonstrates, there is currently a stark trade-off between the use of available stock adjustment data in value-added calculation, and retaining labour productivity observations (this trade-off does not apply to the MFP calculation as stock adjustments and depreciation costs are both sourced from IR10s). As noted earlier, previous New Zealand micro firm data analyses have relied on BAI data, and as such benefit from much lower rates of missing data. Linear interpolation is commonly used for imputation in such cases.

There is the possibility of using IR10 returns that are currently discarded because they fail edit checks. The potential here is not insignificant with 154,164 financial performance returns currently discarded & 60,949 financial position returns discarded in the LBD dataset. Further, for the stock adjustment, there is some potential to patch small holes in longitudinal data since IR10s include both opening and closing stocks. Unfortunately such data is not always consistent across returns (ie, opening stocks in one year do not always match closing stocks in the prior year).

It may also make sense, for some applications, to use AES data in parallel with IR10. Beyond that it is probably sensible to impute stock changes, arguing that the adjustment is minor for most firms. It is another matter as to whether larger scale imputation of IR10 variables is desirable to extend the number of observations of, say, profitability or MFP. The method required would need to be very carefully thought through, given the high rate of imputed data that would eventuate in the dataset. Since the completion of this paper, SNZ have created imputed data for BAI & IR10 making use of a mix of linear interpolation, donor & historical imputation (SNZ 2007).

6. Conclusions

Generally speaking, the breadth of data in the LBD enables significant advances to be made in many areas of microeconomic analysis. This paper represents a first attempt by researchers to exploit some of the potential of the dataset. Given the sheer scale of the data available, we have really only scratched the surface.

The dataset stands out for both its comprehensive coverage of firms operating in the New Zealand economy (having been built primarily around government-administered data collections) and the variety of variables captured. There are distinct advantages from having "full" coverage of firm-level outcomes. For example, this paper has deliberately steered away from tables of means and, instead, focussed on complete distributions, allowing us to see how discussions of average industry labour productivity need to be nuanced by the fact that many firms in "high performance" industries actually have inferior performance to many firms in "low performance" industries (see Bartelsman et al. 2006 for an application of this line of reasoning at the cross-country level). Perhaps more importantly, future industry performance may be related to the dispersion of current firm-level outcomes (eg, if learning effects are more easily transmitted between firms that initially have similar production technologies). The coverage of the LBD makes investigation of such economic models possible.

On the variable side, key strengths of the dataset include the integration of:

- IR10 & AES data enabling the construction of a much wider set of financial performance metrics and, consequently, a more nuanced view of firm success. For example, high profitability & high productivity are not constant companions (Table 7) and decision-makers in firms may be more concerned about the former than the latter;[43]
- LEED variables, enabling better estimation of labour inputs and (aggregate) worker characteristics;
- Customs & IR4 data allowing superior identification of the international linkages of firms (respectively, goods exporting & foreign-ownership/foreign income); and
- A wide variety of sample surveys, expanding the set of research questions that can be tackled with the data.[44]

[43] Policymakers on the other hand naturally focus on the latter. However, any empirical model purporting to explain firm behaviour presumably requires at least a proxy measure of the metric that managers in firms are targeting.

[44] While these surveys have relatively small samples, their location within the LBD provides access to detailed longitudinal performance data, and allows the economic performance of sampled firms to be estimated relative to the full population (eg, Fabling et al. 2008). Further, because of the way business performance surveys are sampled by SNZ, it is possible to construct quite large panels of respondents to multiple surveys (eg, Fabling 2007b).

There is still work to be done with the administrative data to convince the (appropriately) sceptical reader of the validity of derived research outputs. In this regard, the caveats section of this paper has highlighted several key areas for the further development. However, we should try to avoid becoming paralysed by inconsistency in the data. The fact that we have many instances where we have multiple measures of the "same" thing (to the point where, in cases such as wages & salaries, there are four "independent" estimates within the data – AES, IR10, LEED, BOS) is a gift, not a curse, and gives us more choices in our models & our research design. The LBD gives us the advantage of seeing how those choices – choices not available to prior researchers – affect our findings.

At this stage, the analysis presented provides motivation for more work, not definitive answers to questions. This paper is a starting point for more detailed analysis of the dataset. The ongoing research programme is focussed on topics close to the heart of the public policy debate on the economy, particularly improving our understanding of the determinants of New Zealand firm performance, including the impact of government assistance to firms.

The primary weakness of the LBD is the short timespan of the data relative to studies that rely purely on BF/BAI data. Only future years of data production and integration by SNZ can correct this shortcoming. We would encourage other policymakers and researchers to consider whether they have potential uses for the LBD. A larger community of users will deepen our understanding of the New Zealand economy, and can only encourage SNZ to invest further in the development of the LBD.[45]

[45] The IBULDD project's international peer group reviewers advised that increased researcher usage (via the Datalab) was an ideal way to accelerate SNZ's understanding of the data's capabilities and weaknesses, and to improve the data infrastructure for future users (Blanchette et al. 2006).

7. References

Ahn, Sanghoon (2001) “Firm dynamics and productivity growth: A review of micro evidence from OECD countries”, Organisation for Economic Co-operation and Development Economics Department Working Paper 297

Baldwin, John & Wulong Gu (2006) “Competition, firm turnover and productivity growth”, Statistics Canada Economic Analysis Research Paper 42

Bartelsman, Eric, Jonathan Haskel & Ralf Martin (2006) “Distance to Which Frontier? Evidence on Productivity Convergence from International Firm-level Data”, paper presented at Comparative Analysis of Enterprise Data conference 2006, Chicago

Bernard, Andrew & J. Bradford Jensen (1999) “Exporting and productivity”, National Bureau of Economic Research Working Paper 7135

Blanchette, John, Ron Jarmin & Felix Ritchie (2006) “IBULDD project peer group review: Findings and recommendations”, Statistics New Zealand

Carroll, Nick, Dean Hyslop, David Maré, Jason Timmins & Julian Wood (2002) “The turbulent labour market”, paper presented at the NZ Association of Economists’ conference, June 2002

Cox, Val (2006) “IR10 purchases – further analysis”, Statistics New Zealand mimeo

Dixon, Sylvia (2007) “The measurement of labour productivity using the Annual Enterprise Survey and LEED”, Linked Employer-Employee Dataset Research Paper, Statistics New Zealand

Dixon, Sylvia, David Maré & Jason Timmins (2005) “Changes in the distribution of firm size in New Zealand”, Report to the Ministry of Economic Development

Eurostat (2003) “Eurostat manual of recommendations for business registers”, Eurostat

Fabling, Richard (2007a) “The feasibility of producing official statistics from the Longitudinal Business Database”, Statistics New Zealand mimeo

Fabling, Richard (2007b) “Just how innovative are New Zealand firms? Quantifying & relating organisational and marketing innovation to traditional science & technology indicators”, MED occasional paper 07/01

Fabling, Richard, Arthur Grimes and Philip Stevens (2008) “Ask me no questions I tell you no lies: Comparison of qualitative and quantitative firm performance measures”, MED occasional paper 08/04

Fabling, Richard and Lynda Sanderson (2008) “Peter Piper picked a peck of pickled peppers and packed them off to Portugal: Firm-level patterns in merchandise trade”, MED occasional paper 08/03

Grimes, Arthur (1983) “The Reserve Bank real wage-employment relationship: A reply and further results”, New Zealand Economic Papers v17 pp69-78

Law, David, Bob Buckle & Dean Hyslop (2006) “Toward a model of firm productivity dynamics”, New Zealand Treasury Working Paper 06/11

Law, David & Nathan McLellan (2005) “The contribution from firm entry, exit and continuation to labour productivity growth in New Zealand”, New Zealand Treasury Working Paper 05/01

Maré, David & Dean Hyslop (2006) “Worker-firm heterogeneity and matching: An analysis using worker and firm fixed effects estimated from LEED”, Linked Employer-Employee Dataset Research Paper, Statistics New Zealand

Maré, David & Jason Timmins (2006) “Geographic concentration and firm productivity”, Motu Working Paper 06-08

McNaughton, Thomas (2006) “Quality-Adjusted Labour Productivity Measures”, Statistics New Zealand mimeo

MED (various years) “SMEs in New Zealand: Structure and dynamics”, Ministry of Economic Development

OECD (2004) “Understanding economic growth”, Organisation for Economic Co-operation and Development

Scarpetta, Stefano, Philip Hemmings, Thierry Tressel & Jaejoon Woo (2002) “The role of policy and institutions for productivity and firm dynamics: Evidence from micro and industry data”, Organisation for Economic Co-operation and Development Economics Department Working Paper 329

Simmons, Geoff (2002) “Growing Pains: New Zealand Qualitative Evidence on Hurdles to Exporting Growth”, New Zealand Treasury Working Paper 02/10

Smart, Ken & Ross Johnstone (2007) “Year of the exporter: Alternative views of merchandise trade data”, paper presented at the 2007 NZ Association of Economists Annual Conference

Smith, Julie (2006) “Productivity variable pair analysis – 2004 levels”, Statistics New Zealand mimeo

SNZ (2001) “Information about the Business Activity (GST) Indicator”, Statistics New Zealand

SNZ (2007) “Potential Outputs from the Longitudinal Business Database”, Statistics New Zealand

SNZ (various years) “New Zealand business demography statistics”, Statistics New Zealand

Sutton, John (1997) “Gibrat's Legacy”, Journal of Economic Literature v35 No1, pp40-59

8. Tables*

*All results presented here are derived by the authors from the LBD

Table 1 – Counts of economically active, entering & exiting firms by year

Year	Economically active firms	Firms that entered	Firms that exited
2000	442,176	---	---
2001	456,050	51,802	37,928
2002	458,775	50,521	47,796
2003	468,497	56,342	46,620
2004	481,637	58,954	45,814
2005	490,368	56,561	47,830

Table 2 – Patterns of economic activity

Pattern*	**Firms**	**Pattern**	**Firms**	**Pattern**	**Firms**	**Pattern**	**Firms**
Continuous spells							
.....X	47,202	..X...	2,417	.XXX..	4,413	XXXX..	27,682
....X.	2,730	..XX..	3,895	.XXXX.	4,205	XXXXX.	27,021
....XX	47,809	..XXX.	4,737	.XXXXX	31,368	XXXXXX	268,392
...X..	2,307	..XXXX	33,841	X.....	29,824		
...XX.	4,863	.X....	4,039	XX....	35,218		
...XXX	41,671	.XX...	4,442	XXX...	31,480	**TOTAL**	**659,556**
Spells with gaps							
...X.X	646	.X.XXX	579	X..XXX	979	XX..XX	1,204
..X..X	284	.XX..X	307	X.X...	1,147	XX.X..	1,221
..X.X.	204	.XX.X.	239	X.X..X	83	XX.X.X	203
..X.XX	470	.XX.XX	513	X.X.X.	79	XX.XX.	641
..XX.X	519	.XXX.X	528	X.X.XX	133	XX.XXX	1,894
.X...X	245	X....X	860	X.XX..	567	XXX..X	1,486
.X..X.	125	X...X.	343	X.XX.X	110	XXX.X.	1,239
.X..XX	340	X...XX	889	X.XXX.	407	XXX.XX	2,097
.X.X..	249	X..X..	474	X.XXXX	1,628	XXXX.X	2,842
.X.X.X	65	X..X.X	93	XX...X	1,088		
.X.XX.	145	X..XX.	312	XX..X.	540	**TOTAL**	**28,017**

*** An "X" denotes economic activity in the year. For example a firm that experienced a single continuous period of economic activity from 2001-2003 would be represented by ".XXX.."**

Table 3 – Variable coverage rates by source and total employment

Size (total employment)	BAI	IR10	IR4	Positive value-added	Labour prod	MFP
Zero employment	92.2%	51.5%	38.1%	25.6%		
0< employment <=5	95.3%	79.4%	24.7%	60.3%	60.3%	52.7%
5< employment <=20	99.3%	75.2%	64.9%	71.3%	71.3%	67.4%
>20 employment	98.9%	59.7%	75.4%	57.1%	57.1%	53.8%
OVERALL	**94.8%**	**70.8%**	**32.6%**	**51.2%**	**43.9%**	**38.8%**

Table 4 – Variable coverage rates by source and industry

ANZSIC	BAI	IR10	IR4	Pos. value-added	Pos. total employ	Labour prod	MFP
A Agriculture, forestry & fishing	97.1%	73.8%	13.5%	46.8%	80.5%	42.8%	39.8%
B Mining	97.8%	60.9%	56.6%	38.3%	59.2%	31.7%	26.7%
C Manufacturing	96.5%	70.3%	47.4%	56.7%	79.8%	52.5%	47.8%
D Electricity, gas & water	93.9%	60.4%	52.5%	36.7%	47.8%	22.8%	20.2%
E Construction	95.1%	70.8%	30.2%	58.7%	80.2%	54.9%	48.8%
F Wholesale trade	97.2%	67.8%	55.8%	49.3%	67.1%	41.3%	36.7%
G Retail trade	96.2%	71.5%	38.8%	55.2%	79.9%	51.6%	46.2%
H Accom., cafes & restaurants	96.1%	68.5%	40.3%	50.6%	79.0%	47.4%	41.9%
I Transport & storage	95.9%	71.0%	35.6%	52.8%	77.5%	48.3%	41.9%
J Communication services	95.4%	67.8%	13.5%	55.3%	78.6%	52.7%	44.3%
K Finance & insurance	69.8%	71.7%	63.2%	29.0%	47.5%	20.6%	17.2%
L Property & business services	93.6%	68.7%	36.2%	48.8%	54.6%	33.9%	28.3%
N Education	91.3%	73.2%	38.2%	53.9%	78.1%	49.2%	42.6%
O Health & community services	94.6%	74.6%	27.8%	63.6%	86.3%	61.0%	52.6%
P Cultural & rec. services	93.0%	69.2%	29.1%	47.2%	67.1%	39.5%	34.0%
Q Personal & other services	92.9%	73.4%	24.4%	57.2%	82.9%	54.4%	48.0%

Table 5 – Decomposition of count of observations of labour productivity & MFP

			% of total
Total firm-year observations	**2,797,503**		**100.0%**
With positive total employment		1,999,171	71.5%
& BAI sales/purchases data		1,915,663	68.5%
& IR10 stocks data		1,497,202	53.5%
& positive value-added		1,228,322	43.9%
Labour productivity firm-year observations	**1,228,322**		**43.9%**
With positive depreciation costs		1,085,578	38.8%
MFP firm-year observations	**1,085,578**		**38.8%**

Table 6 – Correlation between LFPD and AES postal survey value-added by industry, all years pooled

	Pearson			Spearman rank		
ANZSIC	Level	1yr Δ	4yr Δ	Level	1yr Δ	4yr Δ
A Agriculture, forestry & fishing	0.8442* (1,482 obs)	0.2592* (764 obs)	-0.0105 (104 obs)	0.8639*	0.3609*	0.0336
B Mining	0.9008* (763 obs)	0.3248* (432 obs)	0.0637 (79 obs)	0.9183*	0.4054*	0.1231
C Manufacturing	0.9285* (7,173 obs)	0.3244* (4,354 obs)	0.3247* (836 obs)	0.9516*	0.3821*	0.3405*
D Electricity, gas & water	0.8125* (158 obs)	0.6181* (72 obs)	0.0173 (<20 obs)	0.8243*	0.4923*	-0.0286
E Construction	0.8853* (4,051 obs)	0.3058* (2,189 obs)	0.3090* (318 obs)	0.8976*	0.3837*	0.3108*
F Wholesale trade	0.8485* (4,661 obs)	0.2600* (2,559 obs)	0.2078* (393 obs)	0.8846*	0.2856*	0.2389*
G Retail trade	0.9068* (5,934 obs)	0.2424* (3,539 obs)	0.2060* (619 obs)	0.9242*	0.3148*	0.2283*
H Accom., cafes & restaurants	0.9142* (1,060 obs)	0.2886* (549 obs)	-0.0910 (56 obs)	0.9302*	0.3374*	0.0775
I Transport & storage	0.7747* (2,119 obs)	0.2230* (1,190 obs)	0.2538* (189 obs)	0.7924*	0.3168*	0.2503*
J Communication services	0.8957* (109 obs)	0.4952* (56 obs)	0.3505 (<20 obs)	0.9024*	0.5504*	0.4762
K Finance & insurance	0.6227* (2,503 obs)	0.1907* (1,204 obs)	0.3084* (155 obs)	0.7195*	0.2706*	0.2104*
L Property & business services	0.8503* (8,032 obs)	0.2943* (4,063 obs)	0.3511* (539 obs)	0.8790*	0.4302*	0.3986*
N Education	0.7946 (1,216 obs)	0.2866* (626 obs)	0.2441 (81 obs)	0.7878*	0.3394*	0.2090
O Health & community services	0.8516* (2,315 obs)	0.2611* (1,386 obs)	0.1585 (231 obs)	0.8795*	0.3586*	0.2161*
P Cultural & rec. services	0.8509* (1,708 obs)	0.2801* (807 obs)	-0.0953 (106 obs)	0.8665*	0.3505*	-0.0052
Q Personal & other services	0.8824* (1,197 obs)	0.3579* (676 obs)	0.2268 (85 obs)	0.9143*	0.4002*	0.2691

* A star indicates the correlation coefficient is significantly different from zero at the 1% level

Table 7 – Correlation between measures (levels), all years pooled

	Pearson		Spearman rank	
	labour prod	MFP	labour prod	MFP
MFP	0.7240* (1,085,578 obs)	-	0.6541*	-
profitability	-0.0050* (1,226,574 obs)	0.0071* (1,084,338 obs)	0.1344*	0.1263*

* A star indicates the correlation coefficient is significantly different from zero at the 1% level

Table 8 – Correlation between measures (annual growth), all years pooled

	Pearson				Spearman rank			
	ΔLP	ΔMFP	Δsales	Δemp	ΔLP	ΔMFP	Δsales	Δemp
ΔMFP	0.9673* (643,573 obs)	-	-	-	0.9482*	-	-	-
Δsales	0.5716* (725,380 obs)	0.5054* (643,053 obs)	-	-	0.5419*	0.4790*	-	-
Δemploy	-0.1572* (726,123 obs)	-0.1476* (643,573 obs)	0.2621* (1,320,048 obs)	-	-0.1376*	-0.1356*	0.2655*	-
Δprofitability	-0.0004 (725,380 obs)	-0.0057* (643,053 obs)	0.0024 (1,129,717 obs)	0.0066* (923,064 obs)	0.1627*	0.1802*	0.1943*	0.0088*

* A star indicates the correlation coefficient is significantly different from zero at the 1% level

Table 9 – Labour productivity decile transitions 2000-2005

2000 labour productivity decile	2005 labour productivity decile: 1	2	3	4	5	6	7	8	9	10	
1	1,787	926	602	446	387	318	230	188	219	214	**5,317**
2	1,476	1,532	1,032	714	524	387	364	295	232	261	**6,817**
3	975	1,396	1,461	1,109	861	598	488	363	274	233	**7,758**
4	665	985	1,434	1,489	1,210	898	692	473	377	286	**8,509**
5	562	755	1,103	1,455	1,540	1,222	966	690	466	356	**9,115**
6	415	553	806	1,163	1,510	1,718	1,358	1,039	706	427	**9,695**
7	392	481	585	748	1,185	1,679	1,835	1,508	1,055	581	**10,049**
8	330	349	417	549	824	1,215	1,774	2,147	1,629	788	**10,022**
9	275	315	318	393	549	756	1,148	1,871	2,585	1,550	**9,760**
10	280	253	225	255	285	367	527	848	1,686	3,678	**8,404**
	7,157	**7,545**	**7,983**	**8,321**	**8,875**	**9,158**	**9,382**	**9,422**	**9,229**	**8,374**	**85,446**

Table 10 – Autocorrelation in performance measures (levels and annual growth rates), all years pooled

	Pearson			**Spearman rank**		
	1yr lag	2yr lag	4yr lag	1yr lag	2yr lag	4yr lag
labour prod	0.6305* (726,123 obs)	0.5636* (498,745 obs)	0.4924* (193,491 obs)	0.6829*	0.6188*	0.5474*
MFP	0.7763* (643,573 obs)	0.7346* (442,636 obs)	0.6992* (171,836 obs)	0.8054*	0.7663*	0.7309*
profitability	-0.4426* (1,129,717 obs)	-0.7267* (790,117 obs)	0.0045 (312,338 obs)	0.6979*	0.6253*	0.5359*
∆labour prod	-0.3484* (430,272 obs)	-0.0283* (246,418 obs)	0.0040 (62,472 obs)	-0.2935*	-0.0319*	0.0077
∆MFP	-0.3559* (382,673 obs)	-0.0270* (219,577 obs)	0.0012 (55,714 obs)	-0.2965*	-0.0335*	0.0074
∆sales	-0.1569* (1,268,382 obs)	-0.0349* (840,554 obs)	-0.0053 (229,376 obs)	-0.0559*	-0.0212*	0.0195*
∆employ	-0.0951* (986,683 obs)	-0.0553* (646,152 obs)	-0.0223* (173,694 obs)	-0.0043*	-0.0271*	-0.0011
∆profitability	-0.9984* (717,706 obs)	0.0147* (435,241 obs)	-0.0048 (112,913 obs)	-0.2446*	-0.0251*	0.0095*

* A star indicates the correlation coefficient is significantly different from zero at the 1% level

Table 11 – Decomposition of annual total employment growth from entry & exit

Year	Entering firms	Exiting firms	Change in incumbents	Net change in total employment
2001	74,850	-25,440	-27,170	22,240
2002	72,990	-31,320	-23,450	18,220
2003	78,470	-28,000	-14,410	36,060
2004	76,760	-29,000	-8,320	39,440
2005	74,480	-23,960	-14,000	36,520

*Note: For confidentiality reasons, counts in this table have been randomly-rounded to base 10

9. Figures*

*All results presented here are derived by the authors from the LBD

Figure 1 – Industry labour productivity distribution in 2005

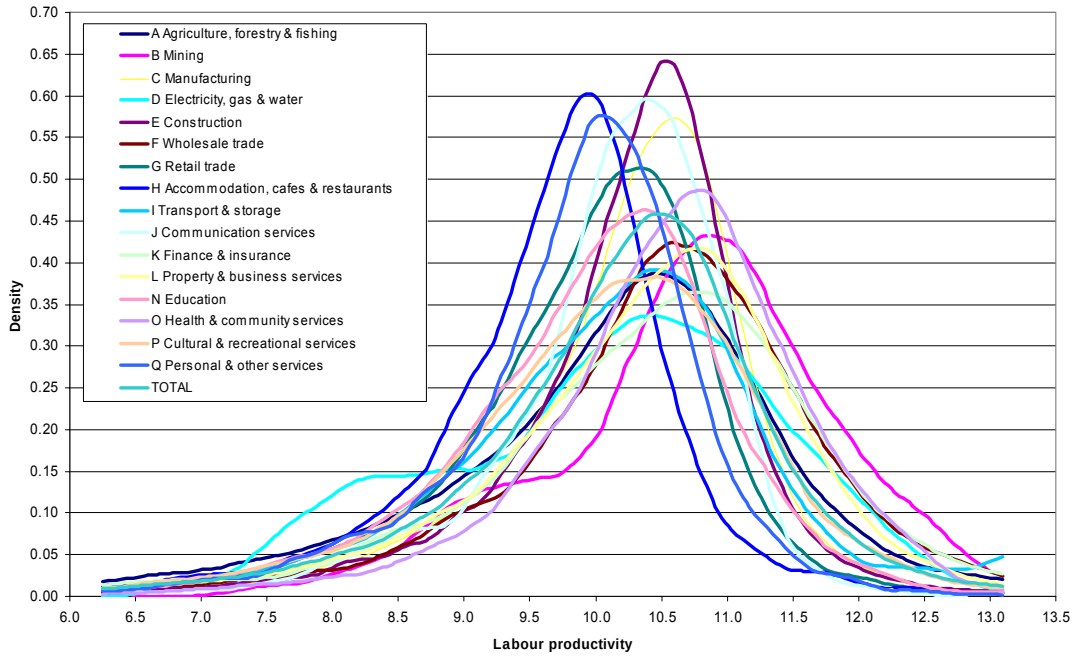


Figure 2 – Industry MFP distribution in 2005

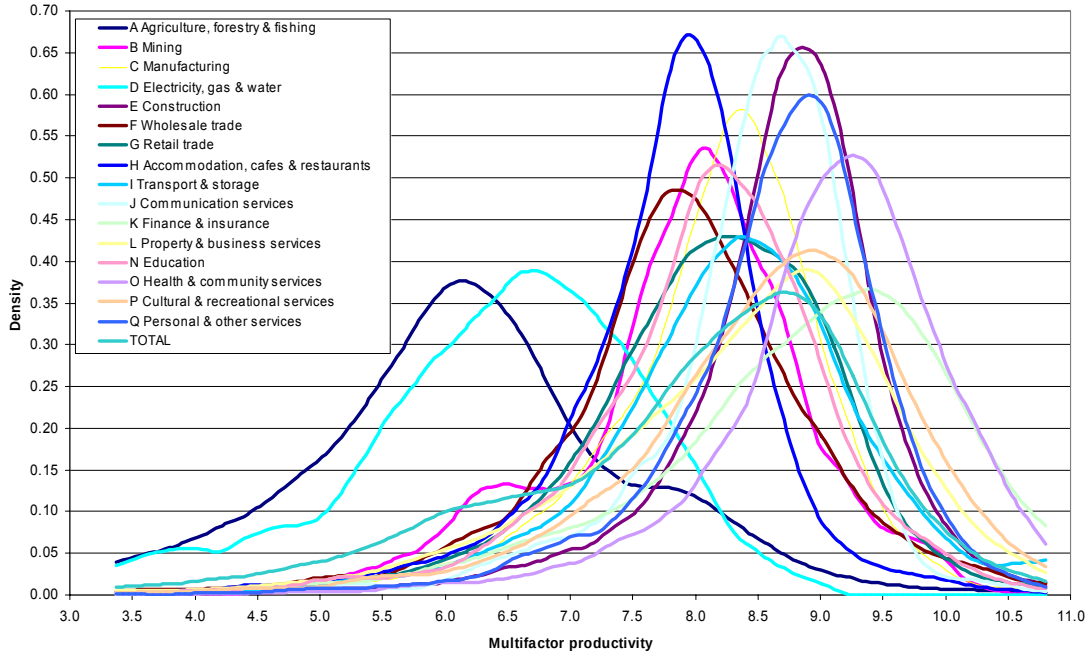


Figure 3 – Industry profitability distribution in 2005

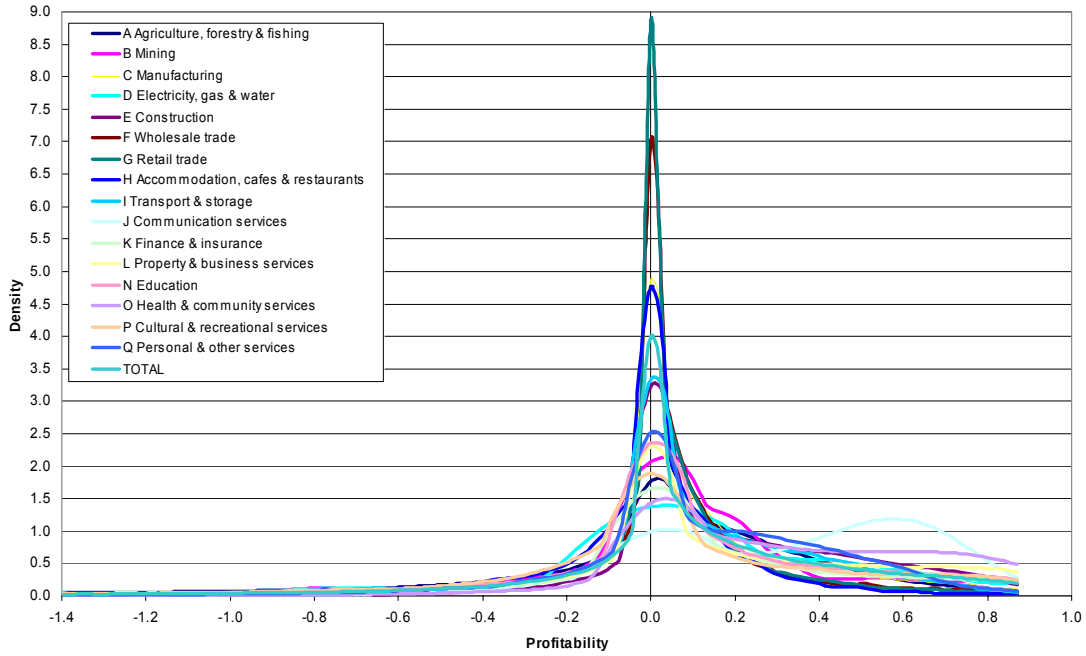


Figure 4 – Aggregate profitability distribution for 2001, 2003 & 2005

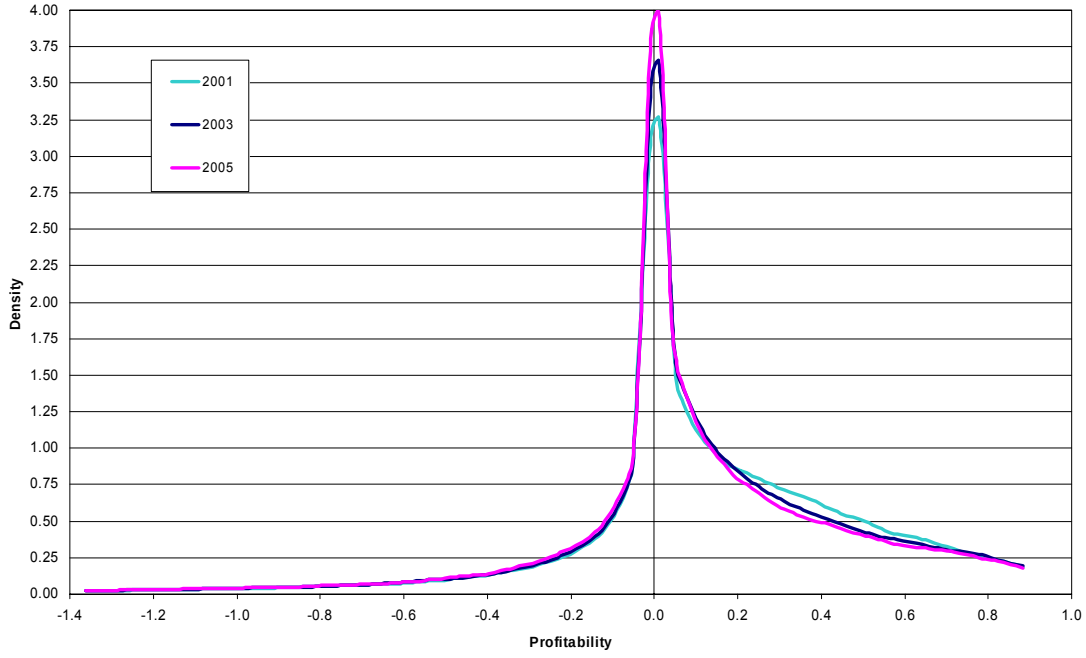


Figure 5 – Aggregate annual labour productivity growth distribution for 2000-01 & 2004-05

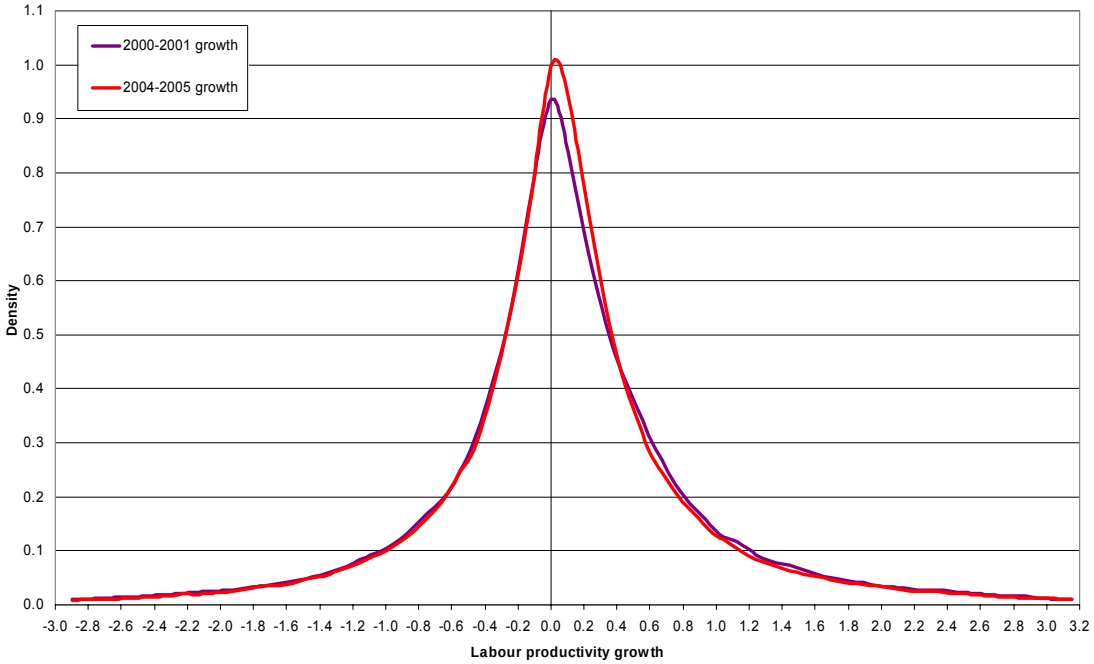


Figure 6 – Entry cohort labour productivity relative to incumbents

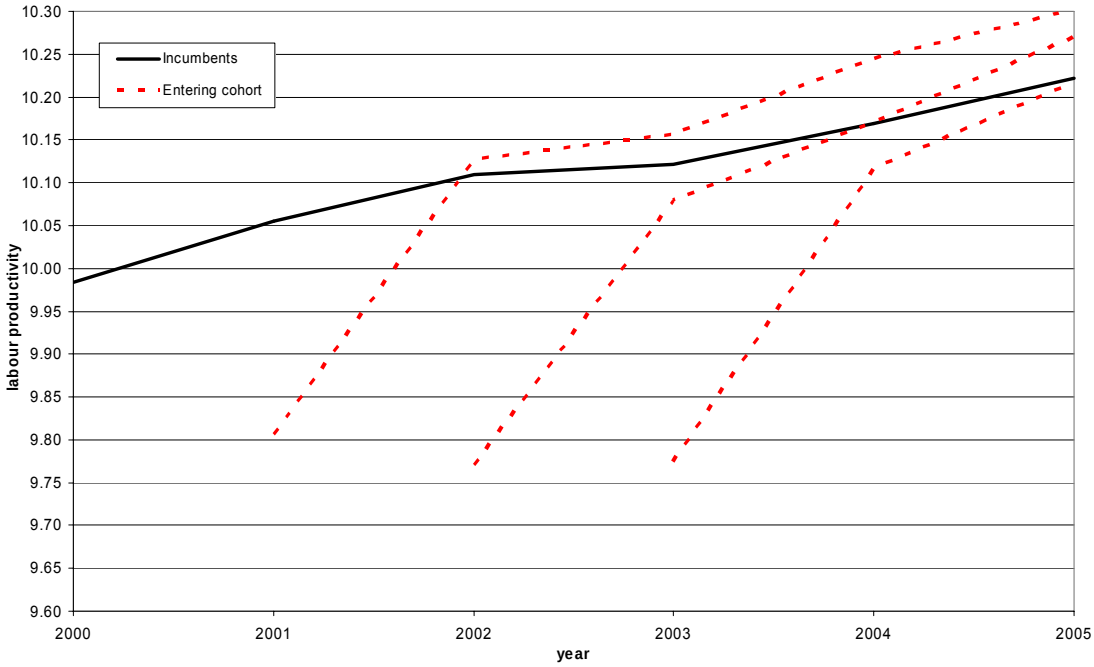


Figure 7 – Exit cohort labour productivity relative to incumbents

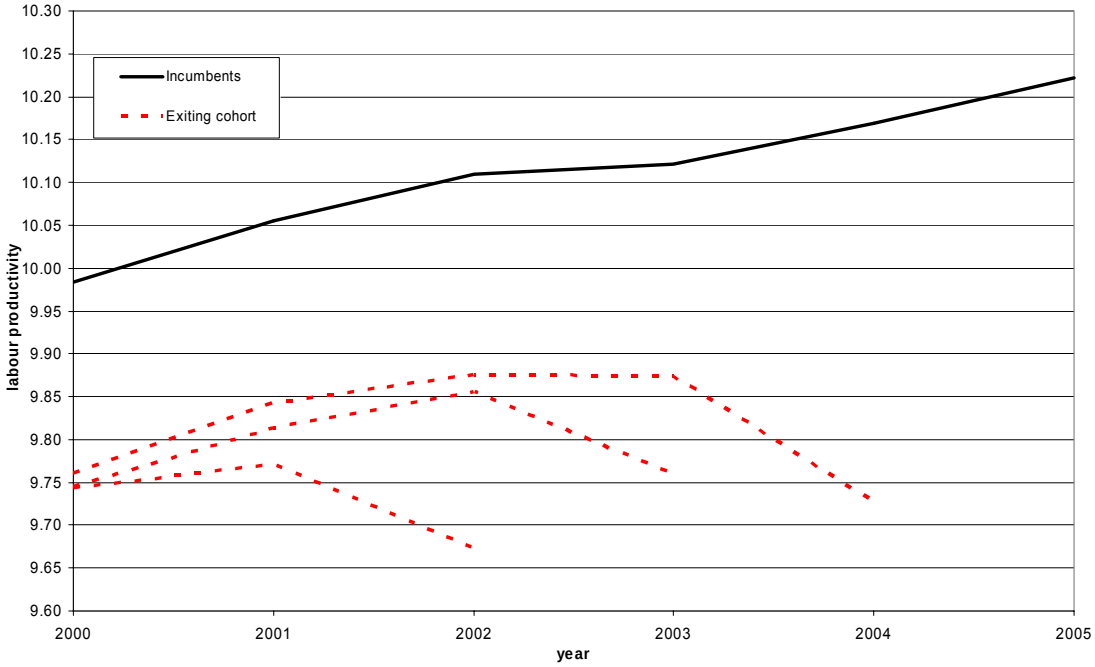


Figure 8 – Histogram of exports as a proportion of total sales for exporting manufacturers in 2005

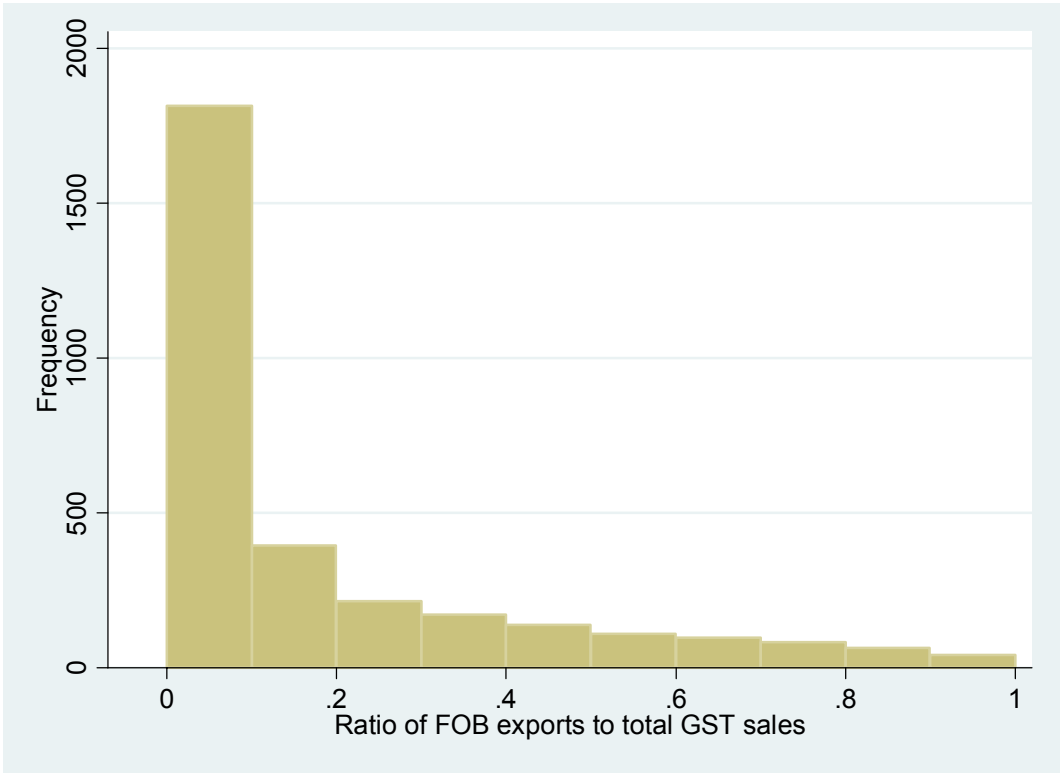


Figure 9 – Average labour productivity by export status and year (manufacturing only)

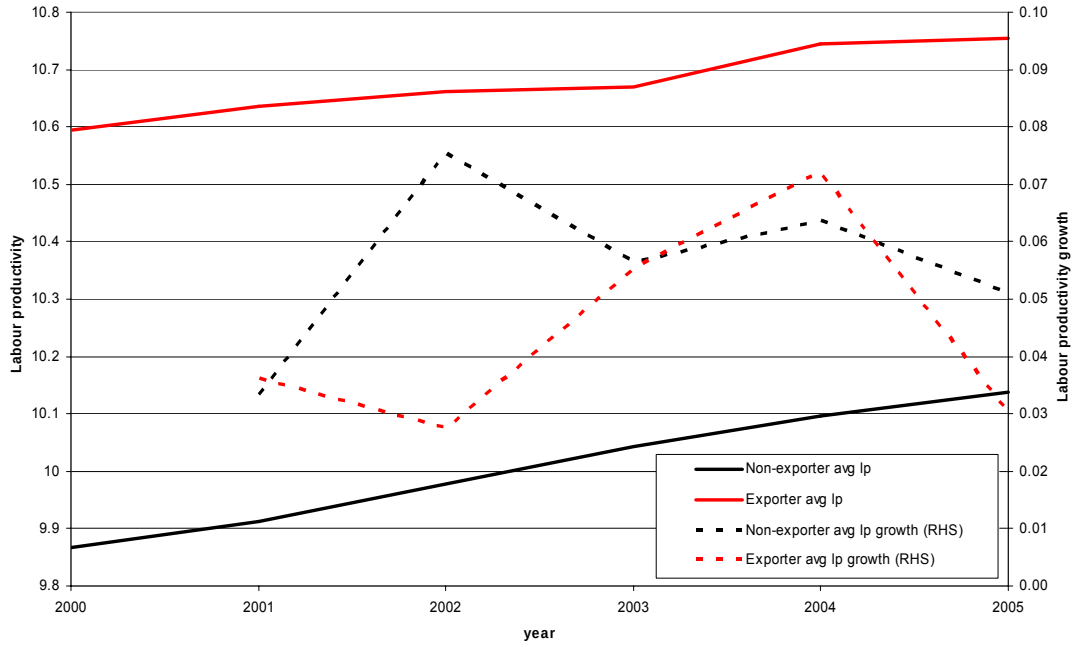


Figure 10 – Annual labour productivity growth distribution by export status (manufacturing only), all years pooled

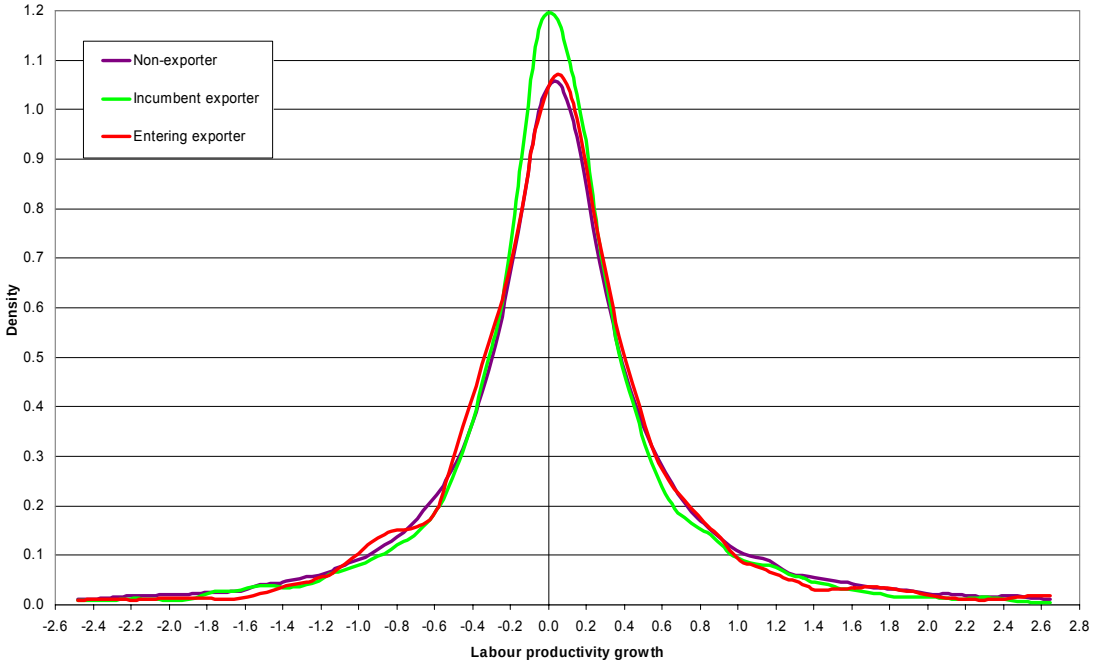


Figure 11 – Labour productivity of foreign-owned vs domestic companies (IR4 filers), all years pooled

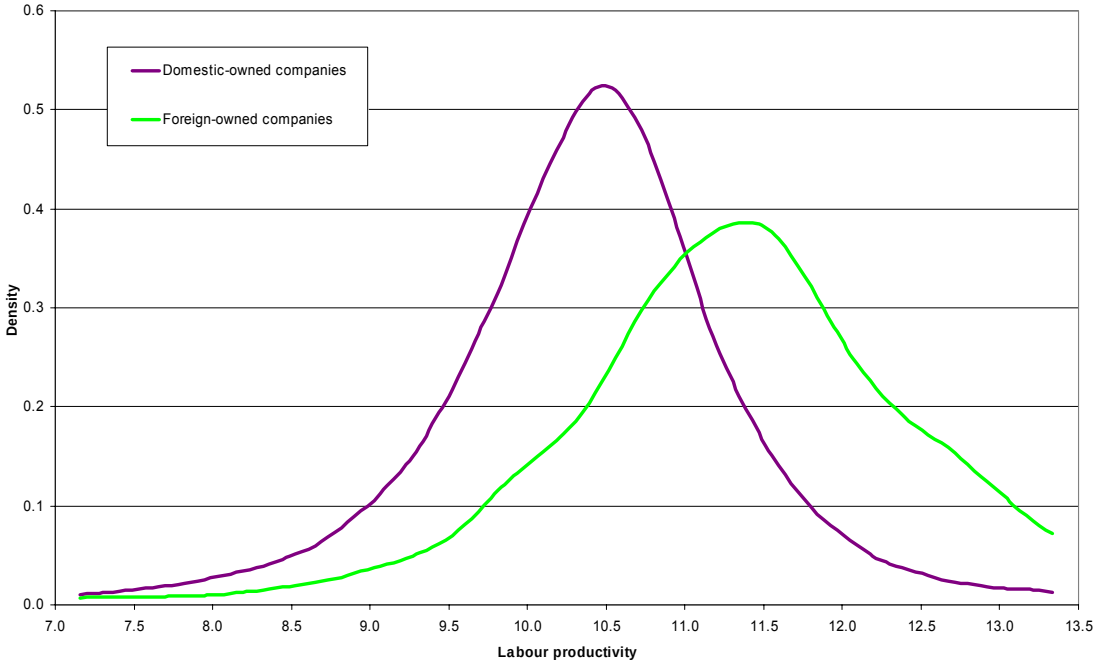
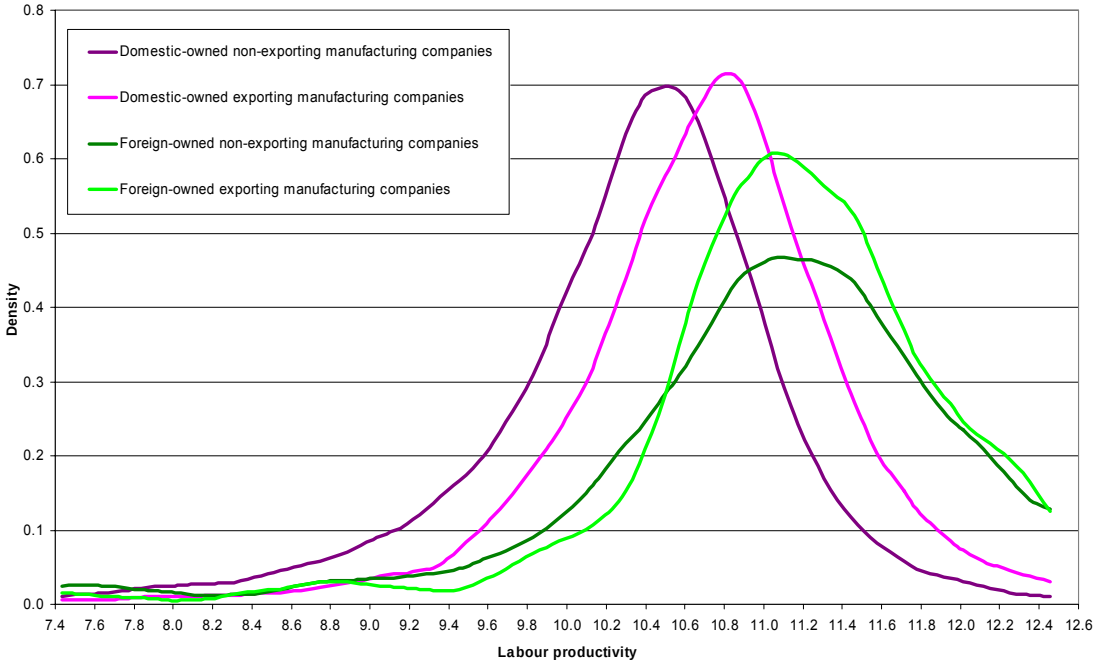


Figure 12 – Labour productivity of manufacturing companies (IR4 filers) by foreign-ownership and export status, all years pooled



10. Appendix A: Multifactor productivity measurement

Table A1 – Multifactor productivity regression coefficients

```
      Source |       SS         df       MS            Number of obs = 1085578
-------------+------------------------------          F(88,1085489) =10697.99
       Model |  1205757.56      88  13701.7904          Prob > F      =  0.0000
    Residual |  1390274.58 1085489  1.28078182          R-squared     =  0.4645
-------------+------------------------------          Adj R-squared =  0.4644
       Total |  2596032.14 1085577  2.39138462          Root MSE      =  1.1317
------------------------------------------------------------------------------
         lva |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             | INDUSTRY CAPITAL COEFFICIENTS
ldep_anz_A01 |   .5081269   .0021634   234.88   0.000     .5038868    .5123671
ldep_an~0234 |   .2921889   .0043273    67.52   0.000     .2837076    .3006702
  ldep_anz_B |   .3141259   .0268367    11.71   0.000     .2615268     .366725
ldep_anz_C21 |   .3474034   .0133106    26.10   0.000     .3213151    .3734917
ldep_anz_C22 |   .2227323   .0112925    19.72   0.000     .2005994    .2448652
ldep_anz_C23 |   .2227869   .0112839    19.74   0.000     .2006709     .244903
ldep_anz_C24 |   .2998311   .0118146    25.38   0.000     .2766748    .3229874
ldep_an~2567 |   .2379428   .0073614    32.32   0.000     .2235146    .2523709
ldep_anz~289 |   .2105203   .0056563    37.22   0.000     .1994343    .2216064
  ldep_anz_D |   .4498305   .0597734     7.53   0.000     .3326766    .5669844
ldep_anz_E41 |   .1949148   .0045093    43.22   0.000     .1860767     .203753
ldep_anz_E42 |   .1951857   .0032908    59.31   0.000     .1887359    .2016355
  ldep_anz_F |   .3187067   .0042921    74.25   0.000     .3102943    .3271192
ldep_anz_G51 |   .2488155    .006201    40.13   0.000     .2366617    .2609692
ldep_anz_G52 |   .2464736   .0041108    59.96   0.000     .2384165    .2545306
ldep_anz_G53 |   .1569383   .0055784    28.13   0.000     .1460049    .1678717
  ldep_anz_H |   .2125509   .0051044    41.64   0.000     .2025465    .2225553
  ldep_anz_I |   .1887827   .0045109    41.85   0.000     .1799415    .1976239
  ldep_anz_J |   .2285929   .0099847    22.89   0.000     .2090233    .2481625
ldep_anz_K73 |    .305399   .0165585    18.44   0.000     .2729449    .3378531
ldep_anz~745 |   .1767957   .0096311    18.36   0.000      .157919    .1956724
ldep_anz_L77 |    .347941    .002957   117.67   0.000     .3421454    .3537367
ldep_anz_L78 |   .2230786   .0024122    92.48   0.000     .2183507    .2278064
 ldep_anz_N  |   .2405153   .0111709    21.53   0.000     .2186207    .2624098
  ldep_anz_O |   .2199395   .0043248    50.86   0.000     .2114631    .2284159
  ldep_anz_P |   .1997699   .0056868    35.13   0.000     .1886239     .210916
ldep_anz_Q95 |   .1398976   .0049718    28.14   0.000      .130153    .1496422
ldep_anz~967 |   .2378468   .0180367    13.19   0.000     .2024954    .2731981
-------------+----------------------------------------------------------------
             | INDUSTRY LABOUR COEFFICIENTS
 lec_anz_A01 |     .57606     .00473   121.79   0.000     .5667894    .5853306
lec_anz~0234 |   .8709443   .0075944   114.68   0.000     .8560596    .8858291
   lec_anz_B |   .7588827   .0468619    16.19   0.000     .6670349    .8507304
 lec_anz_C21 |   .8617243   .0189425    45.49   0.000     .8245976     .898851
 lec_anz_C22 |   1.012209   .0156225    64.79   0.000     .9815893    1.042829
 lec_anz_C23 |   1.060923   .0156663    67.72   0.000     1.030217    1.091628
 lec_anz_C24 |   .9267863   .0184606    50.20   0.000     .8906042    .9629684
lec_anz~2567 |   1.014687   .0106121    95.62   0.000     .9938874    1.035486
lec_anz_C289 |   1.066724   .0082499   129.30   0.000     1.050554    1.082893
   lec_anz_D |   .7628294    .099571     7.66   0.000     .5676736    .9579853
 lec_anz_E41 |   .9918422   .0077712   127.63   0.000     .9766109    1.007073
 lec_anz_E42 |   1.009207   .0057003   177.05   0.000     .9980342    1.020379
   lec_anz_F |   .8924837   .0067673   131.88   0.000     .8792199    .9057474
 lec_anz_G51 |   .8809086   .0081959   107.48   0.000     .8648448    .8969723
 lec_anz_G52 |   .9781721   .0064351   152.01   0.000     .9655595    .9907847
 lec_anz_G53 |   1.051689    .008049   130.66   0.000     1.035914    1.067465
   lec_anz_H |   .9562698   .0069027   138.53   0.000     .9427406    .9697989
   lec_anz_I |   1.154882   .0074723   154.55   0.000     1.140236    1.169527
   lec_anz_J |   .6655936    .017505    38.02   0.000     .6312843    .6999028
 lec_anz_K73 |   .6082279   .0326624    18.62   0.000     .5442108     .672245
lec_anz_K745 |   .9393059   .0176089    53.34   0.000      .904793    .9738188
 lec_anz_L77 |   .5050218   .0070369    71.77   0.000     .4912297    .5188138
 lec_anz_L78 |   .9134526   .0040057   228.04   0.000     .9056016    .9213036
  lec_anz_N  |   .9067592   .0150258    60.35   0.000      .877309    .9362093
   lec_anz_O |   .6520383   .0063366   102.90   0.000     .6396189    .6644578
   lec_anz_P |   .7250736   .0099527    72.85   0.000     .7055667    .7445806
 lec_anz_Q95 |   1.003924   .0092769   108.22   0.000     .9857411    1.022106
lec_anz_Q967 |     .82712   .0326824    25.31   0.000     .7630636    .8911763
-------------+----------------------------------------------------------------
```

	INDUSTRY AVERAGES					
anz_A01	5.951196	.0185742	320.40	0.000	5.914791	5.987601
anz_A0234	7.621776	.0366173	208.15	0.000	7.550007	7.693544
anz_B	7.901176	.2343453	33.72	0.000	7.441867	8.360485
anz_C21	7.048514	.105719	66.67	0.000	6.841308	7.25572
anz_C22	8.121399	.0826707	98.24	0.000	7.959367	8.283431
anz_C23	8.073681	.0872889	92.49	0.000	7.902598	8.244764
anz_C24	7.638774	.0923685	82.70	0.000	7.457735	7.819813
anz_C2567	8.195238	.0566366	144.70	0.000	8.084232	8.306244
anz_C289	8.367611	.0434096	192.76	0.000	8.28253	8.452692
anz_D	6.441179	.5014759	12.84	0.000	5.458303	7.424055
anz_E41	8.6368	.0343038	251.77	0.000	8.569565	8.704034
anz_E42	8.623203	.0250082	344.81	0.000	8.574188	8.672219
anz_F	7.891112	.033136	238.14	0.000	7.826167	7.956058
anz_G51	7.529918	.0477706	157.63	0.000	7.436289	7.623546
anz_G52	7.980345	.0303708	262.76	0.000	7.920819	8.03987
anz_G53	8.814363	.0422972	208.39	0.000	8.731462	8.897264
anz_H	7.770401	.0429196	181.05	0.000	7.68628	7.854521
anz_I	8.463755	.0372089	227.47	0.000	8.390827	8.536683
anz_J	8.430366	.0792843	106.33	0.000	8.274972	8.585761
anz_K73	8.109725	.1350911	60.03	0.000	7.844951	8.374499
anz_K745	8.981705	.0733329	122.48	0.000	8.837975	9.125435
anz_L77	7.415121	.0250666	295.82	0.000	7.365991	7.464251
anz_L78	8.679247	.018553	467.81	0.000	8.642884	8.71561
anz_N	8.09036	.0852021	94.95	0.000	7.923367	8.257353
anz_O	9.054181	.0332357	272.42	0.000	8.98904	9.119322
anz_P	8.565854	.0452598	189.26	0.000	8.477147	8.654562
anz_Q95	8.652761	.038296	225.94	0.000	8.577703	8.72782
anz_Q967	8.251703	.1500409	55.00	0.000	7.957627	8.545778
	YEAR EFFECTS (RELATIVE TO 2005)					
y00	-.2241899	.0037711	-59.45	0.000	-.2315812	-.2167987
y01	-.1554859	.003758	-41.37	0.000	-.1628514	-.1481204
y02	-.0962737	.0037639	-25.58	0.000	-.1036508	-.0888965
y03	-.0884384	.003766	-23.48	0.000	-.0958196	-.0810572
y04	-.0452043	.0037538	-12.04	0.000	-.0525616	-.037847

* ldep refers to log(depreciation costs), lec refers to log(total employment), anz_ refers to industry sub-groups (for example anz_K745 refers to ANZSIC codes starting K74, Insurance, or K75, Services to finance or insurance)

11. Appendix B: Using BAI zero-rated sales as "exports"

Figure B1 – Labour productivity growth distribution by "BAI export status" (manufacturing only), all years pooled