

A rough guide to New Zealand's Longitudinal Business Database

Richard Fabling*

Paper prepared for the plenary session on
Access to enterprise microdata in Asia

Comparative Analysis of Enterprise Data Conference
Tokyo

October 2009

Abstract

Statistics New Zealand's prototype Longitudinal Business Database is a rich resource for understanding the behaviour of New Zealand firms. In this paper we describe the elements of the database, access protocols for researchers, and potential future developments in the linking and availability of business data in New Zealand.

*Reserve Bank of New Zealand (richard.fabling@rbnz.govt.nz); Motu Economic and Public Policy Research. The views expressed in this paper are those of the author and do not necessarily reflect those of the Reserve Bank of New Zealand, or of Motu.

1 Introduction

The prototype Longitudinal Business Database (LBD), maintained by Statistics NZ, is a rich resource for understanding the behaviour and performance of New Zealand firms. The primary selling point of the database is wide coverage administrative sources of financial performance, employment and wage data, coupled with a variety of sample survey and other (notably Customs merchandise trade) data providing a detailed view of firm practices across a broad range of topics. In addition, through the integration of the Business Operations Survey (BOS), researchers have the opportunity to design and integrate entirely new content that can be used with panels of BOS data and other longitudinal information held on the LBD.

The database is a work-in-progress and is continuing to expand as new research needs are identified, and as new data become available. As such this paper can only provide a rough guide to the contents of the database. Section 2 outlines the current components of the LBD, while section 3 covers access and the process for adding data. Finally, section 4 outlines the potential for future developments on both access and content.

2 Components of the LBD

The LBD comprises tax- and survey-based financial data, merchandise and services trade data, a variety of sample surveys on business practices and outcomes, and government programme participation lists (Table 1 provides a summary of the components). The database has expanded organically to meet the needs of agencies whose researchers use the data, and now includes a substantial proportion of the business data that Statistics NZ holds and uses to compile official statistics.

To improve accessibility, each constituent dataset has a “fact table” in which variables are aggregated to a consistent unit of observation and periodicity. Most financial variables are only observed at the enterprise (or tax reporting) level, not at the individual plant, so the enterprise (hereafter, firm) is the common unit of these tables. An annual frequency is imposed on the data by tax-filed financial accounts, working proprietor income declarations, and most sample surveys. Hence, all sub-annual data (goods and services tax, trade, and employment) is annualised to each firm’s financial year and then allocated to the 31st March year-end that has the greatest overlap with the

financial year.¹

Aside from the construction of fact tables, Statistics NZ applies a light touch to the data – the main technologies added specifically for the benefit of LBD users being editing and imputation of tax-filed financial information, and probabilistic matching methods for linking Customs clients and government programme participants to Statistics NZ firm identifiers (Enterprise Numbers).² Where data is transformed, the database usually includes the raw (“load”) data so that researchers can choose to undo any processing that has occurred. For example, while Statistics NZ construct a Customs fact table that includes the aggregate annual merchandise exports of each firm, the database includes the underlying daily shipment level data.

Subsequent subsections outline the contents of the database making note of additional methods for treating data issues that have been developed by the research community.³

2.1 Longitudinal Business Frame and Linked Employer-Employee Data

The Longitudinal Business Frame (LBF) is a monthly “unwinding” of Statistics NZ’s Business Frame – this latter being a set of history tables recording changes in the sampling frame for firms. The LBF contains longitudinal information on – among other things – business name and location, industry, business type, sector, parent-subsidiary relationships, and foreign ownership.

The quality of the LBF’s representation of firm characteristics, and changes in those characteristics, is a function of the maintenance processes for the Business Frame, the ability of respondents to answer survey questions, and the quality of supplementary sources used. Goods and Services Tax (GST) data is used to help maintain the accuracy of the Business Frame (particularly to track the births and deaths of firms and to allocate industry characteristics to small firms) and, consequently, a significance threshold exists at the mandatory GST filing level – previously NZD30K, then NZD40K and now

¹Tax filing rules mean that most firms actually have a 31st March balance date. Variations from that rule are generally industry-specific, eg, the public sector operates to an end of June financial year.

²Statistics NZ also anonymise the data in accordance with the Statistics Act 1975.

³Interested readers can find detailed survey metadata on Statistics NZ’s website www.stats.govt.nz.

NZD50K – below which coverage is limited.⁴ Large economic units are surveyed by Statistics NZ either annually or triennially to maintain the accuracy of the data held.

Because the Business Frame tracks legal entities, firm id continuity can be broken by events that do not imply the exit of a firm. For example, if the partners in a firm decide to reestablish their existing business as a limited liability company that firm may be issued a new firm id despite continuity in location, economic activity and ownership. Fortunately, Statistics NZ puts effort into repairing plant-level ids, making use of continuity of location and employees as measured in the Linked Employer-Employee Data (LEED). Fabling (2007b) quantifies the scope of the issue at the firm level and discusses ways of repairing firm ids using plant-level ids (Permanent Business Numbers), while Fabling and Sanderson (2009a) and Fabling and Grimes (forthcoming) provide methods for constructing “permanent” firm ids.

At the time of writing, the primary source of employment data (LEED) is available for eight (March) financial years 2000-2008 (table 1 includes the temporal span of each data source). LEED is constructed by Statistics NZ from Inland Revenue Department (IRD) tax data, notably Pay-As-You-Earn returns for employees. It is generally assumed by researchers that missing employment data implies zero employees 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.

To protect the confidentiality of individuals, LEED variables have been aggregated to the firm-level before being included in the LBD. Variables available in this manner include counts of employers (on an annual firm-level basis) and employees (on a monthly plant-level basis) with matching firm-level data on income. Summary characteristics of individuals also include gender and age-band 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 corresponding remuneration). This data source is also the basis of an example of a researcher-supplied output that has been added to the LBD – being estimated average worker and firm fixed effects created within LEED (by Maré and Hyslop 2008).

⁴Consistent with the principle of “keeping everything”, tax id-only and Customs id-only records are held on the database where there is filed data that cannot be linked to firm ids. If firms have not been allocated an id by Statistics NZ, they will not have an industry classification.

2.2 GST and the Business Activity Indicator

GST data include information on sales and purchases and is collected on a monthly, bi-monthly or six-monthly basis by IRD, depending on the size of the firm. Statistics NZ 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.⁵

GST in New Zealand has very broad coverage and non-compliance with GST reporting (conditional on filing above the threshold at some point in time) appears very low so that these data are present for almost all employing firms. Certain financial transactions are GST exempt and exported goods are zero-rated. Some researchers have used this latter fact as an indicator of exporting, since the zero-rated component of GST sales is recorded in the data.⁶

2.3 Filed tax returns (IR10 and IR4) and the Annual Enterprise Survey

IR10s are essentially a set of company accounts composed of a profit and loss statement and a balance sheet. Included is information on sales (and other income) and purchases, as well as a detailed breakdown of expenditure including depreciation, research and development costs, and salaries and wages. Balance sheet items include fixed assets (broken down into vehicles; plant and machinery; furniture and fittings; land and buildings; and other), liabilities (current and term), and shareholders' funds.

The Annual Enterprise Survey (AES) is Statistics NZ's primary data source for the production of National Accounts, and as such is the benchmark for estimating 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. IR10s and

⁵GST/BAI data are GST-inclusive whereas most other financial data, including survey responses, are collected on a GST-exclusive basis.

⁶The database contains more detailed export data for goods and services (subsection 2.4). Some income sources unrelated to exports are also zero-rated for GST purposes implying a trade-off between the comprehensive nature of a GST-based indicator of exporting and its likely accuracy.

AES have been used together to construct consistent measures of labour and multi-factor productivity (Fabling and Maré forthcoming).⁷ Maré (2008) provides a clever set of techniques allowing firm-level labour productivity to be used with plant-level location and employment data to examine agglomeration benefits.

IR4s 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.” They also contain a binary foreign-ownership indicator.⁸

2.4 Trade data (Customs and ITSS)

Customs merchandise trade data contains daily shipment-level information over two decades (1988-2008) covering (10-digit Harmonised System) goods, countries of origin and destination, values, volumes, weights, currency of trade, port of entry/exit and mode of transportation. The quality of the match deteriorates (at least in terms of the share of aggregate trade linked) prior to 1996, and some currency variables are only available on a comprehensive basis subsequent to the introduction of mandatory electronic filing (April 2004). The dataset is linked to the LBF initially using tax ids, then probabilistic matching on names and addresses, and finally manually to link any remaining unmatched large-value Customs clients. As with LEED, researchers often assume that the Customs data represents a complete picture of firm-level merchandise trade since the threshold for filing is NZD1,000.

The structure of New Zealand enterprise groups (ie, parent-subsidiary relationships) presents an obstacle to the use of the data, since a large proportion of exported goods pass from manufacturing firms to traders within the group before crossing the border. Fabling and Sanderson (2009a) present a methodology for reallocating trade back to the producing manufacturer.

The International Trade in Services and Royalties Survey (ITSS) provides data on services trade (both exports and imports) by country and service type. The survey runs on a quarterly basis and is an approximate census with a cut-off designed to allow an accurate trade aggregate to be input into

⁷The advantage of using both sources together being that AES has limited coverage of small firms, and IR10s are less likely to be present for firms that are full coverage in AES.

⁸A foreign ownership indicator (percentage of offshore ownership) is also held on the LBF. The advantage of the IR4 indicator is its superior coverage for small companies and potentially more frequent updating for larger firms.

Balance of Payments (BOP) statistics. Two censuses of firms that are likely service exporters (1999 and 2005) appear in the LBD, along with a “birth” survey aimed at ensuring coverage is maintained.⁹ Like the Customs data, the dataset covers a relatively long time period (1996-2008), and is likely to have similar group structure allocation issues, since the target survey respondent is the group-top firm. The integration of this data is still underway at the time of writing, so no research has yet been produced.

2.5 Business Operations Survey and its predecessors (BPS and INN)

The Business Operations Survey (BOS) is an omnibus collection of survey data related to firm behaviour and performance. The population for the survey is all private-for-profit firms with at least a rolling mean employment of six (roughly 35,000 firms). From that population between six to seven thousand useable responses are collected (depending on the year), with the realised response rate always being at least eighty percent.¹⁰

The survey has a modular design (three parts) with Module A collecting annual financial and employment data, and qualitative information on firm performance.¹¹ Module B alternates between collecting innovation statistics (in odd years, using the revised Oslo Manual classification of four types: product, process, marketing and organisational design) and information and communication technology use (in even years, following OECD guidelines). Finally, Module C is open to competitive bidding between government agencies. Topics addressed so far are general business practices (2005 and 2009); employment practices (2006); international engagement (2007); and strategy and skills (2008). In 2010, the primary topic for the module will be price-setting and wage-bargaining sponsored by the Reserve Bank of New Zealand, University of Tasmania, and Victoria University of Wellington.¹²

⁹The ITSS census population is determined by a set of indicators on the Business Frame. These, in turn, are maintained by the Annual Frame Update Survey (firms are asked to report activities over NZD10K) and by monitoring media reports. The history tables for the BOP population indicator set are available in the database.

¹⁰Response is mandatory under the provisions of the Statistics Act for all the surveys in the LBD.

¹¹The quantitative data collected has largely been dropped on the grounds that most of this data is now available from other LBD sources. Outside of the research sphere, the associated reduction in BOS respondent load is probably the greatest achievement of the data integration.

¹²Currently the cost charged by Statistics NZ for the development, testing and administra-

The innovation module and general business practices modules have predecessor surveys also included in the LBD – the Business Practices Survey (BPS 2001) and the Innovation Survey (INN 2003). While there have been changes over time in the way innovation data is collected (reflecting in part changing international standards), there is sufficient overlap in question design and a large enough panel to allow analysis of innovation behaviour between 2001 and 2005 (Fabling 2007a; Maré et al forthcoming).

The BOS sampling strategy includes a top-up element that surveys any respondent to the 2005 survey who would not otherwise be sampled on an ongoing basis – that is, Statistics NZ are running a panel (without replacement) from 2005 onwards. This panel can be linked back usefully to similar qualitative data from 2001. Beyond the explicit panel design, Statistics NZ’s method for random sampling means that there is also substantial repeat sampling of firms across any two years of BOS.¹³

2.6 Other sample surveys (MEUS, BFS and R&D Survey)

The Manufacturing Energy Use Survey (MEUS) 2006 is a precursor to the New Zealand Energy Use Survey (NZEUS). MEUS contains energy use broken down by type (eg, oil, electricity, etc). NZEUS has rolling industry coverage, and will survey the whole economy over a five year period. In 2009, agriculture, forestry, fishing and mining will be surveyed.

The Business Finance Survey (BFS) 2004 was a one-off survey of small businesses, including a mix of quantitative questions about balance sheet structure and qualitative information about the availability of finance. Qualitative questions on the success of firms seeking debt and equity financing have been included in the annual BOS Module A. The 2009 Business Operations Survey includes an expanded set of questions on finance issues, and these additional questions are also largely based on questions originally asked in the BFS.¹⁴

tion of a completely new module is NZD200K (excluding GST of 12.5 percent).

¹³Specifically, Statistics NZ allocate a permanent random number (between zero and one) to a firm when it appears on the Business Frame. Random sampling is then conducted each year by starting at 0.35 on that number line and picking up firms to the right until the appropriate sample fraction is reached. In the absence of entry/exit or redesign of the stratification the same firms would be drawn each year. BOS has been subject to one major redesign (in 2007) to account for revisions to the industry classification system. The survey was dual sampled under the old and new system in that year.

¹⁴These additional finance questions are also scheduled to appear in the 2010 BOS.

The Research and Development (R&D) Survey collects information on (current and capital) expenditure and employment in R&D as well as summary data on funding and the purpose of the research. The survey is conducted every even year and excludes industries that are deemed to be non-R&D performing (eg, wholesale and retail trade). The BOS and IR10s also collect a measure of expenditure on R&D.¹⁵

Due to the way Statistics NZ samples these business surveys, there is a high probability that firms that meet the population criteria for BOS and R&D will be surveyed for both. This is true for any of the sample surveys discussed above, with the exception of the Annual Enterprise Survey which is sampled in a way designed to minimise the overlap with other surveys for smaller firms in the sample.¹⁶

2.7 Government assistance programmes

Lists of firms that have received assistance from government agencies, together with information on the nature of the assistance, have also been probabilistically matched (on contact details) to the dataset to enable evaluation of these schemes (eg, MED 2009). Agencies supplying data are the Foundation for Research, Science and Technology; New Zealand Trade and Enterprise; the Ministry of Tourism; Te Puni Kōkiri; and the Ministry of Social Development. The data includes the period in which the firm received assistance, the specific programme accessed and, where appropriate, an estimate of the value of that aid.

3 Protocols

As can be seen from the breadth of agencies contributing unit-record programme data, the New Zealand legislative environment presents few obstacles to supplying data to Statistics New Zealand, provided there is an end goal of an official statistic. This section covers the legal basis for accessing the LBD, the practical process of applying for research access and for adding new data sources to the database.

¹⁵Changes to the BOS design in 2007 mean that the survey now also collects (in odd years) a broader measure of total product development costs that itemises design and marketing expenditure.

¹⁶This is achieved by AES starting its draw of firms on the Business Frame from the top end of the number line and working left.

3.1 Legal framework

Access to the LBD (and other data) is granted at the discretion of the Government Statistician in accordance with the Statistics Act 1975. Section 37 of the Act conveys the authority to grant access (under controlled conditions) to employees of government departments for research purposes:

Section 37C Disclosure of individual schedules to other Government Departments for bona fide research or statistical purposes

(1) ...the Statistician may disclose individual schedules to any officer of another Government Department solely for bona fide research or statistical purposes pursuant to the functions and duties of that Government Department. ...

Because the LBD includes filed tax returns, the privacy and confidentiality provisions of the Tax Administration Act 1994 also apply to researchers using the data, and the Inland Revenue Department is consulted regarding who accesses the database. Researchers sign declarations to the effect that they have read and understood the relevant provisions of both acts.

For non-departmental researchers, Statistics NZ enables access to the data through secondments – ie, by making the researchers employees of Statistics NZ. To justify a direct secondment, Statistics NZ must be satisfied that the work carried out has benefits to the Official Statistics System.

3.2 Access

The starting point for accessing the LBD is a Datalab application outlining the data required, the research question, the reasons why there is no alternative to using micro data, and the benefits to the Official Statistics System. If approved (either as a government employee or as a secondee) researchers access an anonymised version of the LBD at one of Statistics NZ’s three offices (Wellington, Auckland or Christchurch).

Results of research are subject to confidentiality provisions designed to “protect individual businesses from identification.”¹⁷ The methods used to confidentialise outputs are blunt tools, since they must be effective across an unknowable range of situations. Researchers apply these confidentiality rules to their own results and then must demonstrate to Statistics NZ checkers that the rules have been correctly applied. The process can place substantial

¹⁷Code written by researchers within Statistics NZ’s offices must also be submitted for checking if it is to be taken outside the secure environment.

costs on researchers both in terms of preparing the results for checking and, particularly with new projects or new checkers, in explaining the analysis, outputs and application of the rules.¹⁸ The turn-around time for checking outputs is a maximum of three working days, but is often less in practice.

After results have been written up into a paper, Statistics NZ require the completed work to be returned to them for a final check that the researcher has not inadvertently breached any of the rules through, eg, their description of their method or interpretation of their results. The paper checking process can take up to ten working days. Papers must display a disclaimer explaining how the principles of the legislation have been met and distancing Statistics NZ from the analysis. Further, the LBD must be described as a “prototype” to remind readers that the database is a work-in-progress.

With the support of domestic government agencies and Statistics NZ it has been possible for overseas-based researchers to work with the data though, as indicated above, unit record data is accessible only within New Zealand.¹⁹

Statistics NZ charges for access to the data on a cost recovery basis which may include a fixed fee for any data preparation or linking, and a per-day fee for use of the secure research facilities (the Datalab) and for confidentiality checking.

3.3 Adding data to the LBD

The primary way to add new firm-level data is via a Datalab application explaining the need for research requiring such data to be linked to the LBD.²⁰ Examples of data that have been added through this process include the gov-

¹⁸Statistics NZ may relax the rules where researchers can clearly demonstrate that the standard rule is not necessary to satisfy confidentiality requirements, though it can take some effort to make the case for an exemption and such an exemption is given on an output-by-output basis.

¹⁹There are currently no “general use” business Confidentialised Unit-Record Files (CURFs) in New Zealand. While there are some off-site anonymised and confidentialised datasets, notably unlinked BOS survey data, these are held under tight security for the sole use of the funding government departments. Unconfidentialised unit record data only leaves Statistics NZ in limited circumstances: for BOS, when explicit consent for a follow-up study has been given by the respondent; and where the survey has been run jointly with another agency (for example, the R&D Survey is jointly run with the Ministry of Research, Science and Technology).

²⁰This section covers the integration of additional micro data. There are no general issues with researchers bringing in macro data sources (eg, deflators or trade partner country characteristics) and linking that data directly themselves.

ernment assistance programme data and the trade in services (ITSS) data. Where data integrated in this manner has not been subject to an official first release, then Statistics NZ must publish an official statistic prior to allowing research use.

The other mechanism for adding new data to the LBD is through commissioned BOS content (subsection 2.5 above), since the BOS dataset is currently being updated annually within the database.²¹ The point of difference with the standard process is the ability to design content in the knowledge that additional longitudinal data can be accessed via the LBD (particularly quantitative performance data and other BOS surveys).

Once an agency has successfully applied for additional data to be added to the LBD it becomes accessible to all research users.²² That is, researchers all have access to the same dynamic database.²³

4 Possible developments

There is currently an annual update process, funded by a consortium of government agencies,²⁴ extending the temporal coverage of the LBD. Hopefully in the future, Statistics NZ will gain permanent funding to maintain the LBD. The database can only become more useful as data from the recent economic recession and future recovery is added.

In addition to annual updates, there is still plenty of scope to expand the number of components in the database. There is untapped potential in tax sources particularly from adding sole proprietor and partnership equivalents of the IR4 income tax declaration (the IR3 for sole proprietors and the IR7 for partnerships) and the IR10 profit and loss statement (the IR3B). Statistics NZ hold a number of unlinked business datasets that could form the basis of interesting research projects. While not an exhaustive list, these data include

²¹Technically, the process for the creation of new BOS data is accompanied by a Datalab application enabling access.

²²Subject to the constraint that the work using the data fits with the research question outlined in the researcher's own Datalab application.

²³Archived versions of the data are also accessible to enable replication and journal revisions. A number of the datasets are subject to historical revisions because of, eg, late filed tax returns, tax id-only firms being subsequently included on the Business Frame (hitting the materiality threshold), changes to link tables or firm-level balance dates.

²⁴The Department of Labour; Ministry of Economic Development; Ministry of Foreign Affairs and Trade; Ministry of Research, Science and Technology; Reserve Bank of New Zealand; and the Treasury.

the Quarterly Employment Survey (providing a measure of hours paid); input and output producer price data; Quarterly International Investment Survey (including foreign affiliates information); and a number of industry-specific surveys (eg, Retail and Wholesale Trade Surveys, Accommodation Survey, Agricultural Production Survey,²⁵ and other production survey datasets). Finally, there is the potential to add further firm-level measures of worker characteristics, via an enhanced LEED. Options here could include recent developments in the addition of education data to LEED or, potentially, integration of LEED with five-yearly Census information.

Currently, the integration of any of these data will require government agencies to satisfy the Datalab process outlined above. Alternatively, a more systematic process for determining what the LBD should contain could be developed ending, say, with the database evolving into a central repository for all Statistics NZ-held business data.

Aside from adding to the database, it seems there are opportunities to increase the returns on the current investment by improving accessibility. Those improvements could be made on at least three fronts. Firstly, the Statistics Act could be amended to remove the bias in favour of employees of government departments. It is hard to estimate the potential impact of that restriction. Roughly half the researchers currently accessing the data are not primarily employed by government departments and have their access enabled by direct secondment to Statistics New Zealand or another government department.²⁶ As a rule of thumb, securing access this way is more costly and may, therefore, discourage access. Secondly, all researchers would benefit from a more streamlined Datalab application process – currently perceived as unnecessarily bureaucratic by many who have been through it. Improvements could usefully focus on committing to a time-limited process, shifting emphasis away from trying to assess confidentiality risk at the proposal stage given that sound checking rules apply to outputs, and engaging the applicant more in the assessment process.²⁷ Thirdly, confidentiality rules could be relaxed to allow more of an assessment of risk from first principles – ie, by asking the question, is there a real risk here?²⁸ Such an approach would

²⁵This has a potential partner in the IR3F which is a tax declaration of farming income expenditure, also including detailed information on stocks and farming inputs.

²⁶The Reserve Bank of New Zealand has statutory independence and so is not listed as a department in the relevant legislation.

²⁷More effort on this last point may help overcome a key problem of asymmetric information – the researcher often doesn't know what data Statistics NZ has that may be relevant to their research. Negotiating variations to Datalab agreements to add additional variables can be as time-consuming as the original application process.

²⁸For example, why should the sample size from a randomly sampled survey need to be con-

place more emphasis on the ability of checkers to exercise judgement.

Finally, the LBD is a recent development and there are only a small number of research economists with access, so that very few papers using the database have appeared at international conferences.²⁹ The new cutting edge for firm-level research using Statistics NZ-held data is projects using linked LEED-LBD data. Two approved Datalab applications currently exist looking at the contribution of (LEED-based measures of) worker skills to productivity growth (Hyslop and Maré), and nominal wage rigidity (Fabling).³⁰ Hopefully, as the profile of the LBD improves, more researchers will find use for it – either directly or working collaboratively with New Zealand-based researchers.

confidentialised? Why should researchers not be able to release the maximum and minimum value of the residual produced by estimating a production function? These requirements arise because there are cases where counts and mins/maxs do present confidentiality risks.

²⁹There are approximately ten researchers actively using the data. Two papers were presented at CAED last year in Budapest (Fabling and Grimes forthcoming; Fabling et al forthcoming) and two papers are appearing this year in Tokyo (Fabling and Grimes 2008; Fabling and Sanderson 2009b).

³⁰The latter project also involves the linking (outside of the LBD) of unit record Labour Cost Index data, which is survey information on wage changes for specific jobs within firms.

References

- Fabling, R (2007a), "How innovative are New Zealand firms? Quantifying and relating organisational and marketing innovation to traditional science and technology indicators," in *Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs*, OECD: Paris.
- Fabling, R (2007b), "What chance a longitudinal firm?" Statistics New Zealand mimeo.
- Fabling, R and A Grimes (2008), "Over the hedge? Exporters' optimal and selective hedging choices," *Reserve Bank of New Zealand, Discussion Papers*, DP2008/14.
- Fabling, R and A Grimes (forthcoming), "The "suite" smell of success: Complementary personnel practices and firm performance," *Reserve Bank of New Zealand, Discussion Papers*.
- Fabling, R, A Grimes, and L Sanderson (forthcoming), "Export market choices of New Zealand firms," *Reserve Bank of New Zealand, Discussion Papers*.
- Fabling, R and D Maré (forthcoming), "Production function estimation using New Zealand's Longitudinal Business Database," *Motu Economic and Public Policy Research, Working Paper*.
- Fabling, R and L Sanderson (2009a), "Entrepreneurship and aggregate merchandise trade growth in New Zealand," *Reserve Bank of New Zealand, Discussion Papers*, DP2009/09.
- Fabling, R and L Sanderson (2009b), "Exporting and performance: The impact of destination characteristics on learning effects," paper presented at the 9th Comparative Analysis of Enterprise Data Conference, Tokyo, Japan.
- Maré, D (2008), "Labour productivity in Auckland firms," *Motu Economic and Public Policy Research, Working Paper*, 08-12.
- Maré, D, R Fabling, and S Stillman (forthcoming), "Immigration and innovation," *Motu Economic and Public Policy Research, Working Paper*.
- Maré, D and D Hyslop (2008), "Cyclical earnings variation and the composition of employment," *Statistics New Zealand, LEED Research Paper*.
- MED (2009), "Evaluation of the Growth Services Range: Statistical analysis using firm-based performance data," *Ministry of Economic Development*.

Table 1: Datasets integrated into the Longitudinal Business Database

Acronym	Dataset	Frequency	Availability	Source	Linking	Key population restrictions
LBF	Longitudinal Business Frame	Monthly	Apr99-Sept08	SNZ		Non-employing firms below GST filing threshold
LEED	Linked Employer-Employee Data	Monthly	Apr99-Sept08	IRD	Tax id	
GST	Goods and Services Tax (GST 101 form)	One-, two- or six-monthly	Apr92-Mar08	IRD	Tax id	Mandatory filing threshold
BAI	Business Activity Indicator	Monthly	Apr92-Mar08	IRD	Tax id	Mandatory filing threshold
IR10	Tax-filed financial accounts (IR10)	Financial year	1999-2007	IRD	Tax id	
IR4	Company tax return (IR4)	Financial year	1999-2007	IRD	Tax id	Registered companies only
GAP	Government assistance programmes	Various	1994-2008	Various	Matching	
Customs	Overseas Merchandise Trade Data	Daily	Jan88-May08	Customs	Tax id; matching	Mandatory filing threshold (\$1K); match quality drops pre-1996
ITSS	International Trade in Services and Royalties Survey	Quarterly	96Q2-09Q1	SNZ		Approximate census satisfying BOP materiality threshold
AES	Annual Enterprise Survey	Financial year	1999-2007	SNZ		Stratified random sample; IR10s used for small firms
BOS	Business Operations Survey	Financial year	2005-2007	SNZ		Stratified random sample; 6+RME
BPS	Business Practices Survey	Financial year	2001			Stratified random sample; 6+FTE
INN	Innovation Survey	Financial year	2003	SNZ		Stratified random sample; 10+FTE
R&D	Research and Development Survey	Financial year	1996-2006 (biennially)	SNZ		Stratified random sample; excludes "non-R&D" industries
BFS	Business Finance Survey	Financial year	2004			Stratified random sample; < 100FTE
MEUS	Manufacturing Energy Use Survey	Financial year	2006	SNZ		Stratified random sample; manufacturing; 10+RME

Availability of financial year data relates to "notional" March 31st balance dates (eg, the 1999 year is the financial year ending 31st March 1999 or the financial year that has the greatest overlap with that period if the firm's balance date is not March 31st). FTE=Full-Time Equivalents; RME=Rolling Mean Employment.