Production function estimation using New Zealand's Longitudinal Business Database

Richard Fabling and David C Maré

Motu Working Paper 15-15
Motu Economic and Public Policy Research

September 2015

Author contact details

Richard Fabling
Independent Researcher
richard.fabling@xtra.co.nz

David C Maré
Motu Economic and Public Policy Research Trust
PO Box 24390, Wellington
dave.mare@motu.org.nz

Disclaimer

The results in this paper are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ, Motu Research, or any of the other agencies mentioned above. Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz. The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Motu Economic and Public Policy Research

PO Box 24390
Wellington
New Zealand

Email	info@motu.org.nz
Telephone	+64 4 9394250
Website	www.motu.org.nz

 ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

Abstract

This paper is intended as a resource for researchers using the New Zealand Longitudinal Business Database (LBD) to study the productivity of New Zealand firms. First, it documents the methods used for creating a consistent dataset of production data, combining survey and administrative data sources. Second, it discusses a range of identification and estimation issues that arise when using the data for the estimation of multi-factor productivity. Finally, it demonstrates the value and usefulness of the data by presenting and comparing a range of productivity estimates for a single industry.

JEL codes

C10 [Econometric and statistical methods - general]: D24 [Production; Capital and Total Factor Productivity]

Keywords

Statistics New Zealand Longitudinal Business Database; Production function; multi-factor productivity

Acknowledgements

This paper has been a work-in-progress for many years. It has been amended and updated as we have learnt more, used the data more, and altered our methods to deal with changes in data. Over the years, our work has advanced with assistance from Statistics New Zealand, the Department of Labour, Ministry of Economic Development, Reserve Bank of New Zealand, Ministry of Business Innovation and Employment, New Zealand Treasury, and the Productivity Commission. The work has been completed as part of a research partnership between the Productivity Hub and Motu.

We particularly thank Lynda Sanderson and Dean Hyslop for discussions and comments that have helped us to refine our approach. Thanks also to Adam Jaffe and Trinh Le for comments on an earlier draft of this paper

Errata

This version (16 November 2015) differs from the originally-released version.

- Rows in Table 2 have been labelled correctly
- Table 8a now compares total income on a consistent basis

Contents

Tables

Figures

List of acronyms

AES	Statistics New Zealand’s Annual Enterprise Survey
ACF	Estimation method described in Ackerberg, Caves and Frazer (2006)
ANZSIC	The Australian and New Zealand Standard Industrial Classification
BAI	Statistics New Zealand’s Business Activity Indicator
CME	Commonly measured expenditure (as defined in the text)
EMS	Employer Monthly Schedule (PAYE tax return lodged by employers)
GST	Goods and Services Tax
FE	Fixed Effects
FTE	Full-time Equivalent employment measure
IDI	Statistics New Zealand’s Integrated Data Infrastructure
IR3	Inland Revenue form for individual income tax returns
IR4S	Inland Revenue form for Company shareholders' details
IR7P	Inland Revenue form for Partnership details
IR10	Inland Revenue form for Accounts information and Financial Statement summary
LBD	Statistics New Zealand’s Longitudinal Business Database
LEED	Statistics New Zealand’s linked employer-employee data
NZSIOC	New Zealand Standard Industry Output Categories
OLS	Ordinary Least Square linear regression
PAYE	Pay-as-you-earn system of income tax deductions from wages and salaries
PBN	Permanent Business Number, as defined in Seyb (2003)
PENT	Permanent Enterprise Number, as defined in Fabling (2011)
RLR	Rental, Leasing and Rates component of expenditure
RME	Rolling Mean Employment (annual average of number of employees per month)
QES	Quarterly Employment Survey
VAPW	Value added per worker

1. Introduction

1.1. Objective

This paper is a practical guide for researchers using microdata from Statistics New Zealand's Longitudinal Business Database (*LBD*) to study productivity. It provides an overview of available data sources together with guidance on the selection and processing of variables and firm-level observations. It also contains illustrative examples of productivity estimation, focusing on a single industry – Construction (ANZSIC06 code E11).

1.2. Context: Using microdata for productivity research

Official statistics on productivity provide consistent and reliable estimates of productivity changes, based on index movements in aggregated output and in aggregated productive inputs, and using internationally accepted methods (OECD, 2001). In New Zealand, aggregate statistics are available for the entire measured sector of the economy, and for each of 25 sub-industries within the measured sector (Statistics New Zealand, 2014b). These statistics are essential for monitoring and analysing changes in economic performance over time, and to support economic forecasts.

Firm-level productivity measures are conceptually different from industry or economy-wide measures due to the treatment of intra-industry transactions, and the ability to aggregate productivity measures. Firm-level and aggregate studies can therefore address different questions, and highlight different types and causes of productivity variation. There is, for instance, substantial cross-sectional variation in productivity across firms, which is not observable using aggregate or industry-level official statistics. With wider availability of business microdata, productivity research in recent years has documented and analysed "enormous and persistent measured productivity differences across producers, even within narrowly defined industries" (Syverson, 2011). In the United States, the ratio of productivity between an industry's 90th and 10th percentile plants is 1.92, with the ratio even higher in less developed countries (*ibid*). In New Zealand, the average within-industry ratio among employing firms is 1.84 (Fabling & Sanderson, 2014).[1] Furthermore, inter-firm differences are highly persistent over time, and low-productivity firms are less likely to survive.

Syverson (2011) provides an overview of the wealth of productivity studies that have been made possible by the availability of firm-level microdata, and the range of research questions that can now be addressed. These include: the contribution of firm-level dynamics to aggregate productivity growth; the productivity impact of firm-level characteristics such as managerial practices, firm structure, and input quality and mix; and the influence of factors external to the firm such as competition, local spillovers, and regulation.

New Zealand research in recent years has begun to examine productivity questions using firm-level microdata.[2] There is, of course, still considerable scope for further research. Nolan (2014) discusses a 'forward-looking agenda for research' (*FLARE*) – a list of policy-relevant

---

[1] The reported estimates in Devine et al. (2012) are much larger (an economy-wide index of 7.45) because they use a different measure - the *ratio* of the 90th percentile of estimated mfp to the 10th percentile of estimated *mfp*.
[2] Recent studies include Conway & Zheng(2014), Devine et al. (2012), Doan et al (2014), Fabling & Grimes (2014), Fabling & Sanderson (2013), Grimes et al. (2009), Mai & Warmke (2012), Maré (2008), Maré & Fabling (2013), and Maré & Graham(2013).

research projects identified by the Productivity Hub, a partnership of government agencies that aims to advance research and thinking on productivity issues in New Zealand. The current paper has been prepared within the context of that partnership and, together with the datasets that the methodology generates, provides a general resource that can serve as a foundation for a range of productivity-related research projects.

1.3. Access

Firm-level microdata are not available in the public domain. Access to the data for research purposes is permitted, but under strict provisions that maintain the confidentiality of the data and protect against disclosure of information. Microdata access is governed by Statistics New Zealand, using protocols designed to maintain legislative protections including those under the Statistics Act, and the Income Tax Administration Act. Any requests to access the microdata must be made to Statistics New Zealand and are subject to a stringent approval process for both the research purposes and the researcher(s) involved.[3] Access to productivity data within the *LBD* is currently restricted to selected government agencies.[4] This means that researchers wanting to use the data must have a genuine relationship with a government agency, and be working on topics that the agency wishes to have advanced.

2. General Approach

The focus of this paper is on the processing of data from the *LBD* to support the estimation of multi-factor productivity (*mfp*). At the heart of this exercise is an equation that relates the quantity of inputs used by a firm (i) in a given period (t) to the quantity of outputs produced:

$$Output_{it} = A_{it} * f_{it}(Inputs_{it}) \qquad (1)$$

The function $f(*)$ captures the technology used by the firm, and A_{it} is a measure of *mfp* – the degree to which the firm is more or less efficient at converting inputs into outputs. As written, $f(*)$ is subscripted by firm and year, allowing for the possibility that the firm's technology is idiosyncratic and time-varying. If the changing functional form of the technology were known, it would be possible to estimate *mfp* separately from the technology. In practice, the parameters of the technology are generally estimated jointly with *mfp* so it is necessary to impose some constraints on the technology parameters. It is commonly assumed that $f(*)$ is stable over time, and usually that it is constant for firms in the same industry (j), so $f_{it}(*) = f_j(*)$. In this way, *mfp* is estimated relative to an industry-specific reference technology. A firm with high *mfp* is one that produces more output than other firms in its industry, given the inputs used for production.

---

[3] See http://www.stats.govt.nz/tools_and_services/microdata-access.aspx (Accessed 10 August 2015)

[4] The Government agencies are those that are listed in Schedule 1 of the State Sector Act 1988. The New Zealand Productivity Commission has been granted equivalent status under the New Zealand Productivity Commission Act 2010.

For ease of exposition, we will discuss the case of a Cobb-Douglas production function, which is a relatively simple additive separable function of inputs (indexed by r) when written in logs. The Cobb-Douglas production function has the following form:

$$Y_{it} = A_{it} * \prod_r (X_{it}^r)^{\beta_j^r} \quad (2)$$

or when expressed in logs (with lower-case letters denoting logged variables),

$$y_{it} = a_{it} + \sum_r \beta_j^r \, x_{it}^r \quad (3)$$

Our aim is to derive robust measures of output (Y_{it}) and each factor input (X_{it}^r) in order to estimate the parameters (β_j^r) of the production function and *mfp* (a_{it}).

2.1. Inputs and outputs

The simplicity of equation (2) belies the practical challenges of defining and measuring a firm's inputs and outputs. The task is straightforward if the firm produces a single quantifiable output and uses a small number of distinct quantifiable inputs. In practice, researchers commonly use a single index that summarises the quantity of output, even when the firm produces multiple outputs. Similarly, diverse inputs are generally classified into a few composite input types – typically under the headings of capital, labour, and intermediate inputs.[5] We follow this convention and develop measures of gross output (Y_{it}) and three classes of inputs: capital (K_{it}), labour (L_{it}) and intermediate inputs (M_{it}). This is sufficient to support two common forms of production function estimation – gross output and value added.[6] Using lower case letters to denote logs, the Cobb-Douglas version of these forms is as follows:

$$Gross\ Output: y_{it} = a_{it} + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it}$$
$$Value\ Added: va_{it} = ln(Y_{it} - M_{it}) = a_{it} + \beta_k k_{it} + \beta_l l_{it} \quad (4)$$

The measures shown in the above equations denote the quantities of inputs and outputs. As is commonly the case when estimating productivity using microdata, what we observe are firm revenues and expenditures, with the exception of labour inputs, for which employment counts are available. In order to convert these accounting measures into indexes of quantities, we must deflate gross revenue, capital value or expenditure, and expenditure on intermediate inputs by appropriate price indices. Each component must be deflated separately. In the absence of firm-level measures of output or input prices, this is done at the industry level, using available price deflators.

The modelling of output quantities as a function of the quantities of inputs should not be confused with the accounting decomposition of the (dollar) value of gross output into different components of expenditure and surplus. The following diagram summarises the accounting distinctions that lie behind the calculation of input and output quantities.[7] The economic relationships in equation 4 are relationships between *quantity* of output or value added and the

[5] Alternative breakdowns are used where studies are focused on particular types of input: eg: intangible capital; high-skilled v low-skilled labour; land; ICT.

[6] See OECD (2001) or Cobbold (2003) for a discussion of the advantages and disadvantages of each approach.

[7] In practice, the accounting equality is not always satisfied in the data.

quantities of inputs associated with the expenditures on intermediate consumption, compensation of employees, and capital services.

Figure 1: Accounting components of the value of gross output

<table>
<tr><td></td><td></td><td colspan="4">Value Added</td></tr>
<tr><td>Gross output =</td><td>Intermediate Consumption</td><td>+ Compensation of employees</td><td>+ Capital Charges</td><td>+ Indirect taxes</td><td>+Net Operating Surplus</td></tr>
<tr><td></td><td></td><td></td><td></td><td colspan="2">Gross Operating Surplus</td></tr>
<tr><td></td><td colspan="4">Total Expenditure</td><td></td></tr>
</table>

2.2. The challenges of *mfp* estimation

Estimating *mfp* requires not only the measurement of inputs and output, but also choices about the functional form of the relationship between inputs and output, as captured by a production function or cost function. Identification issues also arise, related to our ability to estimate the parameters of the production or cost function from observed variation in inputs and outputs.

When people measure or talk about *mfp*, they are generally interested in somewhat broader concepts of 'productivity'. In general, 'productivity' refers to a firm's efficiency in converting inputs into output. In the case of a single input, this idea translates naturally into measures such as labour productivity (eg: value added per worker). It is a relatively simple measurement exercise to evaluate whether a firm has a high or low level of labour productivity. Concluding that two firms with equal labour productivity are equally efficient in converting inputs into outputs relies on an implicit benchmark that output should increase proportionately with labour input.

Regression-based approaches to measuring productivity answer a slightly different question. They relax the assumption that output should increase linearly with labour input, and use a benchmark relationship between output and inputs, captured by a 'production function', $f(*)$ as in equation 1. In equation 1, *mfp* is captured by the term A_{it}.

By addressing this slightly different question, regression-based approaches sacrifice the ability to make direct (ratio) comparisons of productivity between firms with different levels of inputs. Comparing the productivity of firms with identical inputs can still be done directly on the basis of the ratio of outputs to inputs. However, statements about the relative productivity of firms with different levels of labour input depend on the assumed or estimated relationship between inputs and outputs. Each firm is judged by comparing its ratio of output to inputs with that of a benchmark calculated for firms with the same level of inputs. It is this normalised ratio that is used for comparison between firms with different input levels.

Multi-factor productivity is thus essentially a relative concept. Efficiency is measured relative to a benchmark level of efficiency, as captured by the production function. The benchmark is a function of inputs and can be set in a number of ways, including as the average predicted output or as the maximum (frontier) level of output. Because the measure is relative, it depends crucially on how the benchmark is constructed. In the production function context, this means that the measure of *mfp* will depend on what range of inputs are controlled for, and what function of them is used to calculate the benchmark.

If we were to model outputs as a function of labour inputs alone, we may find that firms with higher stocks of capital are more productive. We may conclude that capital is a source of productivity advantage. Alternatively, we could define benchmark productivity based on the joint use of labour *and* capital, in which case capital would be included with labour in the production function. Whether or not capital-intensive firms are identified as relatively productive will depend on the modelled benchmark relationship between capital intensity and output.

A further consequence of the relative nature of *mfp* as a measure is that it is not possible to compare the relative efficiency of two groups of firms that operate with completely different technologies. The efficiency of each group is measured relative to its own benchmark. Where industries operate with different technologies, for instance, and hence different production functions, *mfp* is meaningful only for comparisons within the same industry – inter-industry comparisons are not possible (eg: comparing *mfp* for manufacturing firms with that of retail industry firms).

These somewhat trivial examples serve to highlight some of the key challenges of *mfp* estimation. A measure of *mfp* is meaningful only when interpreted in the context of which inputs are controlled for, and how the benchmark is constructed for how inputs are expected to be transformed into output. The challenge of interpreting variation in *mfp* is magnified by the likelihood that not all relevant inputs or differences in technologies are controlled for, or even observed. It is in this sense that *mfp* is "a measure of our ignorance" (Abramovitz, 1956).

In this section, we discuss different approaches to establishing regression-based benchmarks for firm efficiency, through choosing different production functions, and different approaches to controlling for unobserved differences between firms, either econometrically, or by adding additional control variables. In general, the more inputs or characteristics are introduced, or the more refined the econometric specification, the more we load inter-firm differences into the benchmark, and the less is attributed to *mfp*.

2.2.1. Functional form

The shape of the production function is generally estimated from variation across firms and over time in the relationship between inputs and outputs. The function is generally estimated using a specific parametric function, chosen from among a wide range of suitable candidate functions.[8]

There are few *a priori* restrictions on the functional forms of the relationship between inputs (x_i) and output (y), other than that we would expect that an increase in all inputs would not reduce output. Restrictions are, however, generally imposed to support the interpretation of findings in the context of economic production theory. Obtaining estimates of marginal products, elasticities of substitution, and returns to scale, for instance, require the use of functions that are at least differentiable. Such measures are commonly used to summarise key features of production technologies, and for use as parameters in simulation and modelling. Different functional forms place different restrictions on the size, sign and behaviour of key measures. We focus here on a linear function, and three of the most commonly used functional forms: Cobb-Douglas (CD), Constant Elasticity of Substitution (CES), and Translog (TL). Equations for the four functions are shown in the second column of Table 1. The remaining columns show some key differences between them.

[8] There is a large literature using cost functions (the dual of the production function) to estimate production parameters (see Jorgenson, 1986).

The third column shows the implied marginal product of input x_i derived from each function. The formulae illustrate the different ways that marginal productivity is allowed to vary with input quantities. Using a linear production function, the marginal productivity of input x_i is modelled as a constant – not depending on the quantity of any inputs. Marginal productivity could be positive or negative. With Cobb-Douglas and CES functions, the marginal productivity of an input can vary with the quantity of other inputs. We generally expect that the marginal productivity of x_i will decrease as the quantity of x_i increases and increase with the quantity of other inputs. The signs are, however, not restricted, though the sign does not change as the mix of inputs changes. This is not true for the translog function. The flexible functional form allows both the size and the sign of marginal productivity to vary with the quantity of any input.

These differences are obviously important if estimates are going to be used to learn about the marginal productivity of different factors when using different input mixes, since the answers will be constrained by the function used. They may be less important if the focus of analysis is *mfp*, though it is an empirical question as to whether *mfp* estimates differ across specifications (see section 6).

The elasticity of substitution (column 4 of Table 1) is a summary measure of how the mix of optimally chosen inputs changes in response to changes in relative prices. In equilibrium, relative prices are equated to relative marginal productivities, so the elasticity reflects the shape of the production function. For any pair of inputs x_v and x_w,

$$\sigma_{vw} = \frac{dln(x_v/x_w)}{dln(f_w/f_v)} \; where \; f_z = \frac{\partial y}{\partial x_z} \qquad (5)$$

The functions shown in Table 1 differ markedly in the constraints that they impose on factor substitutability. With the linear production function, any pair of factors will be used together only if prices are exactly equal to their (fixed) marginal productivities. The elasticity of substitution is thus infinite, as any change in relative prices will lead to the use of only one of the factors. Both the CD and CES functions restrict σ_{vw} to be constant, and the same for all pairs of inputs. In the case of the CD function, the elasticity is restricted to equal 1, so that a percentage increase in the price of a factor is accompanied by an equal percentage decrease in the quantity used, with factor cost shares therefore remaining constant. The CES function does not constrain the size or sign of σ_{vw}, but maintains the restriction that it is a constant and common for all pairs of inputs. Of the functions shown, it is only the translog function that allows σ_{vw} to vary across input pairs.[9]

The flexibility of the translog function is thus able to provide estimates that are more informative about the interrelated demands for factors of production, and the nature of differences between different technologies used by different firms or over time. It also admits greater variation in factor shares over time.

The implications of the different functions for factor shares are determined by the degree of substitutability between inputs. The differences are also evident from the formulae for marginal products. Under perfect competition, the real (deflated by output price) price of an

[9] The nested CES function is an alternative that allows for variation in substitution elasticities across different pairs of inputs but restricts the structure of substitutability. For instance, output may be modelled as a CES function of two inputs, X and L_u (unskilled labour), $Y = A\left[\alpha X^{\psi} + (1-\alpha)L_s^{\theta}\right]^{\frac{1}{\psi}}$ where X is itself a CES function of capital (K) and skilled labour (L_s): $X = B\left[\beta K^{\theta} + (1-\beta)L_s^{\theta}\right]^{\frac{1}{\theta}}$.

input is equated to its marginal product, so a factor's cost share (θ_x) is the marginal product of the factor multiplied by the ratio of the quantity of input to the quantity of output:

$$\theta_x = \frac{p_x x}{p_y y} = \frac{\partial y}{\partial x}\left(\frac{x}{y}\right) \tag{6}$$

This cost share equals the output elasticity of the factor $\left(\frac{\partial lny}{\partial lnx}\right)$. It is evident from the formula for the marginal product that, for the CD function, the factor share is a constant (β_i). For CES, the factor share changes monotonically with changes in any one input, and for the translog, the changes are less constrained. More flexible functional forms will thus be more informative for research questions related to factor cost shares.

Table 1: Some properties of common functions

Function	Equation	Marginal Product $\left(\frac{\partial y}{\partial x_r}\right)$	Elasticity of Substitution	Homogeneity & Homotheticity	Number of parameters, given n inputs
Linear	$y_i = \alpha + \sum_r \beta_r x_i^r$	β_r	$\sigma = \infty$	Homogeneous of degree one if $\alpha = 0$ Homothetic	$n+1$
CD: Cobb-Douglas	$y_i = \alpha \prod_r (x_i^r)^{\beta_r}$ Or, equivalently, $\ln(y_i) = \ln(\alpha) + \sum_r \beta_r \ln(x_i^r)$	$[\beta_r]\left(\frac{y}{x_i^r}\right)$	$\sigma = 1$	Homogeneous of degree $(\sum_i \beta_i)$ Homothetic	$n+1$
CES: Constant Elasticity of Substitution	$y_i = \left(\alpha + \sum_r \beta_r (x_i^r)^{-\rho}\right)^{-\frac{\nu}{\rho}}$	$\beta_r \nu (x_i^r)^{-1-\rho}\left(\alpha + \sum_i \beta_i (x_i^r)^{-\rho}\right)^{\frac{-\nu-\rho}{\rho}}$ $= \left[\nu \frac{\beta_r (x_i^r)^{-\rho}}{(\alpha + \sum_s \beta_s (x_i^s)^{-\rho})}\right]\left(\frac{y}{x_i^r}\right)$	$\sigma = \frac{1}{1+\rho}$	Homogeneous of degree ν if $\alpha = 0$ Homothetic	$n+3$
TL: Translog	$\ln(y_i)$ $= \alpha + \sum_r \beta_r(\ln(x_i^r))$ $+ \sum_r \sum_s \delta_{rs}(\ln(x_i^r))(\ln(x_i^s))$	$\left[\beta_r + 2\sum_s \delta_{rs}(\ln(x_i^s))\right]\left(\frac{y}{x_i^r}\right)$	Non-constant	Homogeneous of degree $(\sum_r \beta_r)$ if $\sum_r \delta_{rs} = 0$ $for\ all$ r. Homothetic if $\sum_r \delta_{rs} = 0$ $for\ all$ r.	$\frac{(n+1)(n+2)}{2}$

Source: Adapted from Griffin et al. (1987, Table 1)

Finally, the functions differ in the constraints that they impose on the way that output expands as all inputs are increased proportionally. A production function is said to be homogeneous of degree d if a doubling of all inputs raises output by a factor of 2^d. In the case of $d=1$ (i.e. homogeneous of degree 1), the production function is said to exhibit constant returns to scale, since a doubling of all inputs leads to a doubling of output. None of the functions in Table 1 impose homogeneity but all allow homogeneity as a (testable) restriction (column 5). The linear, CD and CES functions do, however constrain the way that the mix of inputs is permitted to change as output increases. Specifically, inputs are constrained to increase proportionally when output increases, as long as relative prices do not change. This property is called homotheticity, and ensures that expansion paths take the form of rays from the origin. Only the TL function allows for non-homothetic production, whereby the factor mix can change as relative prices change, and to test for homotheticity. Allowing for non-homothetic production is important for some research questions, such as distinguishing whether a changing factor mix reflects a change in technology or is a consequence of output changes with a non-homothetic technology.

The Table 1 functions are only a small subset of functional forms that have been used for production function or cost function estimation. Griffin et al. (1987) presents and discusses a broader range of functions that have been used to examine different aspects of production technologies or to relax particular constraints in commonly used functions.

2.2.2. Estimation, identification and endogeneity

Although the CD function is relatively restrictive, it is perhaps the most commonly used function for production function estimation. It is parsimonious – requiring only $n+1$ parameters to be estimated, and can be estimated (in log form) using linear regression. The CES function has two additional parameters and is non-linear in parameters, requiring alternative estimation approaches such as non-linear least squares. The need for non-linear estimation has in the past made computation more difficult but with currently available software, non-linear models can be easily estimated. The flexible functional form of the TL specification comes at a cost – the number of parameters to be estimated increases rapidly, with the square of the number of factors. Furthermore, the additional parameters capture the curvature of the production function, which is often more difficult to estimate precisely due to collinearity among the variables (Fuss, McFadden, & Mundlak, 1978). While such collinearity is less problematic if the aim is prediction, it is clearly a disadvantage if there is interest in interpreting particular production parameters.

Generally, more flexible functional forms are preferable if the data can support sufficiently precise estimates. Restrictions can in some cases be tested, as for the TL function, which nests the CD, or the CES function, which nests CD and linear functions (Berndt & Khaled, 1979; Giannakas, Tran, & Tzouvelekas, 2003).

Whatever functional form is chosen to capture production technologies, the estimation of production parameters from observed data such as is available in the *LBD* poses some fundamental identification challenges. Griliches & Mairesse (1998) provide a clear exposition of the issues and discussion of econometric treatments. They present a simplified version of an estimating production function equation:

$$y = \alpha z + \beta x + (a + e + \varepsilon) \tag{7}$$

where y is the log of output, z is the log of fixed factors of production (such as capital, which cannot readily be altered in the short term), and x is the log of variable factors of production (such as labour, which can be adjusted in the short term). The term in brackets is a residual. The essence of the econometric identification problem is that this residual, or components of it, are likely to be correlated with the quantity of one or more factor inputs, making estimates of α and β obtained from ordinary least squares (*OLS*) regression biased and inconsistent.

To explain the source of correlation, Griliches & Mairesse distinguish three components of the residual term. The first, a, is a disturbance that is observed by the firm but not by the econometrician. Firms choose the quantity of variable inputs, x, taking account of a. A positive value of a is expected to increase a firm's demand for variable inputs, generating a correlation between x and a. The second error component, e is also observed by the firm but does not affect the choice of inputs in the short term, though if it changes the firm's expectation of long-term productivity, it may affect input choices in subsequent periods. Finally, the third error component, ε, is 'the econometrician's problem' and reflects the net error from measurement, data collection and computational procedures – none of which affects firms' choices.

The first component, a, is thus the primary source of simultaneity between inputs and disturbances. The appropriate econometric response to this problem depends on the nature of the component. Unfortunately, there are many different sources of simultaneity, including omitted (time-invariant) attributes or inputs, time-varying productivity shocks, survival bias, and cross-firm differences in prices. Different econometric specifications can control for particular types of simultaneity but no single specification can control fully for all potential types. In general terms, econometric responses have taken one of three forms: 1) transform variables to eliminate a; 2) use only that part of variation in x that is orthogonal to a (an instrumental variables approach); and 3) include a proxy for a (a control function approach).

Various forms of difference transformations have the potential to remove the simultaneity bias if the elements of a are primarily unobserved attributes of a firm that are relatively fixed over time, such as the quality of inputs, management effectiveness, unobserved inputs such as land, intangibles or intellectual property, or errors due to functional form mis-specification. These include mean differencing (firm fixed effects) and time differencing over short or long timespans. Such differencing will eliminate any variables that are time-invariant.[10] As discussed by Griliches & Mairesse (1998), such differencing has the disadvantage of exacerbating other econometric problems. In particular, it reduces the variance of relatively stable inputs, such as capital, and magnifies any downward bias from errors in variables or measurement error. Difference transformations also fail to control for the effect of time-varying components of a, or of e, which may be transmitted to subsequent factor choice.

[10] Controlling for firm fixed effects may remove bias from the estimated coefficients but the fixed effects themselves should be included in estimated *mfp*.

The second econometric response to simultaneity is to identify coefficients from the variation in *x* that is orthogonal to a. This is an 'instrumental variables' approach, and requires variables that do not directly affect production and are correlated with factor choices but are uncorrelated with a. Candidate instruments include factor prices, though these are seldom available at the firm level, and lagged values of inputs. The validity of instruments will depend on the way that a and e are transmitted to factor choices.

The third econometric response, which has been widely used in recent years, is to include in the estimating equation a 'control function' – a variable or function that is correlated with a. Fixed effect estimation can be thought of as a control function approach – the inclusion of a firm-specific intercept controls for a to the extent that a itself is a time-invariant firm-specific component. A series of recent papers have taken the approach of including a control function that captures variation in 'proxy variables', which the firm adjusts in response to a or e but which do not directly affect output (Ackerberg et al., 2006; Levinsohn & Petrin, 2003; Olley & Pakes, 1996; Wooldridge, 2009). Olley & Pakes use investment as a proxy variable whereas Levinsohn & Petrin use materials inputs. In either case, the proxy variable (q) is assumed to be chosen taking into account the quantity of fixed inputs and the realisation of the disturbances $(\omega = a + e)$. Under some mild assumptions, the function $q = q(\omega, z)$ can be inverted to yield an expression for ω in terms of the observable proxy and fixed factors $\omega = \omega(q, z)$. The estimating equation then becomes

$$y = \alpha z + \beta x + \omega(q, z) + \varepsilon \quad (8)$$

The form of $\omega(q, z)$ is unknown, so it is approximated by a flexible polynomial in q and *z*. Because z is included both in the polynomial and in the estimating equation, an extra assumption is needed to allow α to be estimated. A sufficient condition is that the disturbance evolves over time as a Markov process: $\omega_{it} = E[\omega_t|\omega_{it-1}] + \xi_{it}$.[11] Within this framework, Olley & Pakes use the correlation between $\omega(q, z)$ and firm survival to also control for estimation bias arising from the loss of less productive firms from their panel. As in the case of fixed effects, the firm-specific component $\omega(q, z)$ should be included in estimated *mfp*.

The methods described so far have been focused on the estimation of production function parameters, with *mfp* estimates obtained from regression residuals. There is a wide range of alternative approaches tailored to address specific questions about productivity and production technologies, or to accommodate particular features of market structure or data availability. A prominent stream of the literature deals with estimation in the presence of imperfect competition in either input or output markets – generally also dealing with the simultaneity challenges discussed above.

With imperfect competition, input and output prices can be endogenous, and may thus vary across firms. This is of particular concern when estimating productivity using datasets that lack firm-specific price and quantity measures, such as the *LBD*. The *LBD* dataset contains deflated revenue and deflated intermediate expenditures rather than output and intermediate quantities. Any estimation of production function parameters using these measures will confound parameters of the underlying technology with variation in prices.

[11] This is a simplified exposition. See Ackerberg et al. (2006) and Wooldridge (2009) for a fuller discussion of the conditions under which α can be identified in equation 8, the potential use of additional moment conditions, and approaches to estimation.

To the extent that the ability to gain high prices reflects product quality or innovative outputs, this may be a meaningful measure of productivity in some contexts. It is thus plausible to incorporate market power into the measure of productivity, and rely on measures of 'revenue productivity' rather than physical productivity.

The alternative is to control for price variation when modelling the relationship between measured inputs and outputs. In this way, estimated *mfp* reflects technical efficiency, and the structure of imperfect competition is captured by the production function. There is no single best way of incorporating price effects in productivity estimation. There is a range of approaches, which vary in the assumptions that are made, the methods used, and in the production parameters that they attempt to estimate.[12] For instance, Klette (1999) and Klette & Griliches (1996) allow for imperfect competition and derive estimates of demand and scale elasticities and the heterogeneity of market power, scale effects and productivity. Martin (2010) extends this approach, identifying firm-specific markups, and thus estimating the distribution of both productivity and market power across firms. Grieco et al. (2013) deal with the case of input price dispersion arising from imperfectly competitive input markets. New approaches and extensions to existing approaches are constantly being developed, to deal with new identification concerns, new research questions, and new data.

3. LBD Sources

3.1. Identifying firms

Financial information in the *LBD* is available at the enterprise level. Where an enterprise operates in more than one location ('plants'), there is employment information for each plant. Enterprises are assigned a unique identifier (*enterprise_nbr*) that is intended to identify each enterprise longitudinally. We improve the longitudinal tracking of enterprises, using information on longitudinal identifiers developed at the plant level (*pbn_nbr*) to create 'permanent enterprise numbers' (*PENT*s), following Fabling (2011). Each enterprise is associated with a unique *PENT*, so enterprise-level production data can be aggregated to the level of *PENT*s.

Each *PENT* is assigned to a unique industry, based on the industry of the enterprise that accounts for the greatest share of employee-months. The identification of predominant industry is based on New Zealand Standard Industry Output Categories (*NZSIOC*), which are groupings of ANZSIC06 industries. Appendix 1 shows the 54 industry groups that we use.[13] This level of industry detail is chosen because Producer Price Index (*PPI*) series are available for each of these groups.

[12] Key dimensions of difference include the use of first order conditions, the use of parametric or semi-parametric production functions, assumptions about the timing of input choices, choice of orthogonality conditions, and the choice of which relationships to use as estimating equations.

[13] Prior to 2006, industry data (eg: Annual Enterprise Survey, Business Operations Survey) were classified using the *ANZSIC96* industry coding. *ANZSIC06* codes have been allocated retrospectively to *LBD* data from earlier years, imputing from available *ANZSIC96* codes. *ANZSIC06* codes are missing for a small number of *PENTs* with neither industry code. Earlier versions of the productivity code used *ANZSIC96* industry codes grouped into the 2-digit industries shown in Appendix Table 2.

3.2. Survey data - Annual Enterprise Survey

The benchmark data source for production function estimation is the Annual Enterprise Survey (*AES*). It is a primary source of information for the estimation of national accounts, and is designed to provide annual data for financial performance and financial position by broad industry groups. Because the concepts and measures used in the survey are designed for the purposes of production measurement, the data are the most appropriate for use in production function estimation.

The term '*AES'* is used by Statistics New Zealand to refer to two different things: first, to a postal sample survey of firms; [14] and second, to a compiled dataset of business information that includes data from the sample survey, but also includes data from administrative sources.[15] Over time, the size of the sample survey has declined as the reliance on alternative data sources has grown, reflecting Statistics New Zealand's commitment to reducing respondent burden. In 2012, the *AES* sample survey collected survey responses from 3.6% of units[16] in the target population, down from 14.9% in 1997. Sampled units are, however, predominantly large units, which collectively accounted for around half of total employment. The postal survey is stratified by industry and size, with full coverage of enterprises in the largest-size stratum within each industry.

For production function estimation, we select only *AES* postal survey responses from the *AES* dataset, (identified in the *LBD* table *fact_aes_enterprise_year* by *data_srce_code='Postout'*). The imputation and modelling that Statistics New Zealand carry out for *AES* data from other sources, while appropriate for cross-sectional or industry-level estimation, are an unreliable basis for longitudinal firm-level analysis. Data are also imputed for around 25% of the AES postal survey sample, and we discard AES data for these firms. Non-imputed AES responses are identified using the variable *fact_aes_enterprise_year.unit_status_code*, for which we retain observations with codes of C (clean) or S (suppressed warnings).[17] The imputation flag is missing in the year 2000, so it is not possible to identify which data are imputed. In order to maintain consistent data quality, we do not use that year for productivity estimation.

Finally, we drop AES postal responses that are derived from form types that do not provide the necessary production information or are out of scope. These are forms with values of *load_aes_header.questionnaire_id_code* of the following types: AL (consolidated return); NP (not-for-profit); CP (commercial property); or miscellaneous unknown form types (XX, NULL, MV).

---

[14] Appendix 2(a) contains a copy of an *AES* postal survey for the *Manufacturing and Wholesale* industry. The form itself varies across broad industry groupings, reflecting a desire to consistently capture the most important components of output and inputs on and industry-by-industry basis. Further information on the *AES* is available under the heading of 'Data quality' on the webpage for the latest release – accessible at http://www.stats.govt.nz/browse_for_stats/businesses/business_finance/annual-enterprise-survey-info-releases.aspx, and at http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/36809771-984d-4e6b-89a1-576f2118b05b (accessed 8 August 2015).

[15] For the 2012 *AES* release, data were compiled from the sample survey; business financial data from Inland Revenue (*IR10*); central government data from the Treasury's Crown Financial Information System; superannuation from the New Zealand Companies Office; local government data from Statistics New Zealand's local authority statistics; and not-for profit data from the Charities Commission.

[16] 'Units' are *KAUs* (kind-of-activity units) rather than enterprises. The sample is selected by enterprise and each enterprise may contain more than one *KAU*.

[17] Values of this variable that identify discarded observations are E (error); M (imputed due to poor data quality); R (imputed due to non-response); W (imputed financial position data).

AES financial data are recorded in thousands of dollars. We multiply these by 1,000 so that they are in dollars, consistent with the recording of financial information in administrative tax forms.

3.2.1. AES discontinuities

Statistics New Zealand notes that *AES* data are not necessarily comparable from year to year: "AES is designed to measure industry levels for a given year. Incremental improvements in measurement, sample design, classification, and data collection may influence the inter-period movements, particularly over longer time periods."[18]

The AES questionnaire was redesigned in 1999, and again in 2009. Accompanying the 2009 redesign, there was a substantial decrease in the number of units sampled, linked to an increase in the use of administrative (*IR10*) data – a change that was extended further in 2010, 2011, and 2012. In 2010, the sampling strategy was made more efficient, and there were 'enhancements to editing and imputation processes'. The sample was boosted in 2006 to support the transfer of industrial classification from *ANZSIC96* to *ANZSIC06*. In 2007, the sample was 're-optimised' to suit design and publication on *ANZSIC06*.[19]

3.3. Tax data - IR10

Where we do not have useable *AES* survey information for a firm, we draw information from firm-level administrative data, in the form of summaries of accounts information provided for tax purposes to Inland Revenue in the form of *IR10* tax forms (see Appendix 2(b)). The *IR10* form has two components, covering firms' financial performance (profit and loss), and financial position (balance sheet). We restrict attention to *IR10* forms that have passed some rudimentary quality checks for consistency and completeness of both the financial performance (front-page) and financial position (back-page) responses.[20]

Whereas the *AES* postal survey uses specialised questionnaires (of approximately 16 pages) to collect appropriately measured financial information from each of 28 industry groups, the *IR10* information uses a single 2-page form. Not surprisingly, the data provided in the *IR10* form are neither as detailed as, nor entirely consistent with, the information obtained from the *AES* postal survey. Crucially for the purposes of productivity estimation, assets and expenses are grouped differently across the two data sources, requiring some imputation and modelling at the firm-level, as described below.

3.3.1. *IR10* discontinuities

Because the *IR10* data are collected for administrative (tax) purposes, the concepts and definitions used reflect the needs and the legislative parameters of the tax system. When tax

[18] See http://www.stats.govt.nz/browse_for_stats/businesses/business_finance/annual-enterprise-survey-info-releases.aspx (accessed 8 August 2015).
[19] ibid.
[20] Passing of edit checks is identified in the table *fact_i10_enterprise_year* by (*fp_edit_status_code*='p' and *bp_edit_status_code*='p')

laws change, definitions of variables such as taxable income, depreciation, deductible expenses, and asset valuations may also change. In 2011, for instance, there was a change in the tax treatment of depreciation, which was reflected in substantial declines in estimates of economic depreciation. Statistics New Zealand report that the legislative change contributed to the large (5.9%) decline in aggregate depreciation, and to an even more pronounced decline (37.4%) for the rental, hiring, and real estate services industry.[21] It is unclear whether these changes in reported depreciation were also reflected in responses to the *AES* postal survey, though this is likely.

Over time, as the reliance on administrative data has increased, there has also been greater reliance on modelling key economic variables that are not well captured in the administrative sources, as the following quotation from Statistics New Zealand, which accompanied the release of the 2012 *AES* results, indicates:

> ***Use of administrative data and its effect on published variables***
>
> *Our main administrative data source (Inland Revenue's IR10) is the primary source for capturing the agriculture, forestry, and fishing division (ANZSIC06 division A) in AES. In 2012 we used more administrative data for other industries as well. IR10 data does not provide direct estimates of additions and disposals of fixed assets, so we use modelling to calculate these. The modelling of IR10 data is currently under review, so additions and disposals of fixed assets have been suppressed from the 'all industries' table, all agricultural industries, and the accommodation industry tables in this release. Our increased use of administrative data in 2012 has also caused discontinuity in shareholders' funds and owners' equity in the repairs and maintenance industry, and the accommodation industry.*

A revised *IR10* form was introduced in 2012/13, with changes reflecting both administrative and statistical data needs.[22] It is not yet clear what impact, if any, this will have on series continuity or on the methodology outlined in this paper.

3.4. Tax data – GST

Another relevant source of administrative data comes from firms' goods and services tax (*GST*) returns. *GST* is a tax on the consumption of almost all goods and services in New Zealand. All businesses that conduct taxable activity and have turnover larger than a minimum turnover threshold are required to register for *GST* and file periodic (monthly, bi-monthly or six-monthly) returns.[23] This source therefore provides a measure of sales and purchases for a broad subset of New Zealand firms.

Some early research using the *LBD* relied on net *GST* payments as a measure of value added, or to identify export activity (Fabling, Grimes, Sanderson, & Stevens, 2008; Maré & Timmins, 2006). Subsequent research has used a combination of *AES* and *IR10* data, as outlined below.

---

[21] For details, see http://www.stats.govt.nz/browse_for_stats/businesses/business_finance/AnnualEnterpriseSurvey_HOTP12/Data%20Quality.aspx (accessed 8 August 2015).

[22] For details, see http://www.ird.govt.nz/technical-tax/general-articles/ga-revised-ir10-summary.html (accessed 8 August 2015).

[23] Some goods and services are GST exempt or zero-rated, including GST-exempt supplies. GST exempt supplies include residential dwelling rentals, financial services and donated goods and services sold by non-profit. Zero-rated goods and services are taxable activities that are taxed at a rate of 0%. The majority of zero-rated goods and services are exported for use outside of New Zealand.

This choice reflects the availability of more detailed and internally consistent financial performance and financial position information available from *AES* and *IR10* sources.

Our only use of *GST* data for productivity estimation is to help detect firm activity by year, in order to identify transitions into and out of operation.

3.4.1. GST discontinuities

The main change over time in the *GST* data is the change in the turnover threshold below which firms are not required to register for *GST*. The threshold for actual or expected annual turnover increased from \$30,000 to \$40,000 on 1 October 2000, and since 1 April 2009, has been \$60,000. These changes mean that the coverage of *GST* returns has probably declined over time, though firms below these thresholds are required to register for *GST* if they charge *GST* on their sales or wish to claim back *GST* paid on purchases.

3.5. Tax data - EMS (PAYE) payroll information

The primary source of employment information used to measure labour input is the Employer Monthly Schedule (*EMS*) – a monthly tax return filed by employers summarising the monthly wage and salary payments made to each of their employees, and the 'pay-as-you-earn' (*PAYE*) income tax deductions made. These data are used to identify the number of employees in each firm in each month. Employees who receive self-employed income as well as wages and salaries from the same firm are classified as working proprietors and not counted as employees.

The *EMS* data are sourced from the Integrated Data Infrastructure (*IDI*) under the *clean_IR* schema.[24] We aggregate employment and working proprietor information to the *PENT*-year level. The derived measures of labour input that are included in the *LBD* productivity dataset are measured in full-time-equivalent (*FTE*) units, using an algorithm described in section 4.5 and documented more fully in Fabling & Maré (2015).[25] *PENT*-year measures of average monthly *FTE* and headcount employment, and annual working proprietor counts are available within the *LBD* in the *ibuldd_research_datalab* database as table *pent_year_L_IDI_yyyymmdd*, which also includes information on firms that are not included in the productivity dataset.

3.5.1. EMS discontinuities

There have been no discontinuities in the coverage or reporting requirements for PAYE tax, which is included from April 1999. Prior to 2015, the *LBD* contained summary annual information for each plant (*pbn_nbr*) and enterprise (*enterprise_nbr*) derived from *EMS* data contained in Statistics New Zealand's linked employer employee data (*LEED*). These tables have been discontinued. We prefer to rely on the Fabling & Maré (2015) employment measures rather than use the Longitudinal Business Frame measure of rolling mean employment (*RME*) that is used by Statistics New Zealand for tasks such as sample selection. Our preference reflects that fact that the Fabling & Maré measures are adjusted to some extent for part-time or

[24] Researchers with access to the *LBD* who have a legitimate research use for the *EMS* data should satisfy the criteria for access to the linked *LBD-IDI* data used here.

[25] This algorithm includes adjustments that reduce an individual's labour input to be less than full time if they hold multiple jobs, earn less than the monthly minimum wage, or receive some forms of non-work benefits.

part-month employment, and treat self-employment more appropriately for productivity estimation (see section 4.5).

3.6. Other data sources

The derived firm-level labour input measures rely indirectly on annual tax returns to identify working proprietors within each firm. These are identified from a combination of income tax returns (*IR3*: earnings not taxed at source); partnership tax returns (*IR7P*: partners receiving distributions); and company tax returns (*IR4S*: shareholder salaries).

4. Defining populations and variables

In compiling the *LBD* data into a form that can be used for productivity estimation, we follow a few general principles. Where possible, we use data from non-imputed responses to *AES* postal survey questionnaires. Where such data are not available, we use data from *IR10* accounts information provided to Inland Revenue for tax purposes. We restrict attention to industries in the measured sector, identified by Statistics New Zealand (2014a) as "industries that mainly contain enterprises that are market producers. This means they sell their products for economically significant prices that affect the quantity that consumers are willing to purchase". The measured sector restrictions are summarised in Appendix Table 1. Consistent with the market producer definition, we also restrict to private-for-profit businesses.[26]

When using *AES* data, we can simply aggregate variables that are necessary for calculating productivity components. Flows are summed across all contributing enterprises in a *PENT* in a year. Opening (closing) stocks are based on the enterprise that is the first (last) that is associated with the *PENT* during the year.[27] Because *AES* also contains questions on depreciation, additions, disposals, and gains and losses from sale, it is possible to reconstruct an opening book value for each asset class.[28] We prioritise the use of lagged closing book values over derived opening book values for consistency with *IR10* treatment, and because derived opening book values are sometimes implausible (eg, negative values).

When using *IR10* data, additional manipulation is needed to create measures that are as comparable as possible to the *AES* measures.

[26] Private-for-profit firms are identified by excluding those firms where business_type>6 and which are not a State Owned Enterprise or a Local Authority Trading Enterprise.
[27] In practice, there are so few *PENT*s with multiple enterprises reporting *AES* data that a summation is used. The technically correct aggregation approach is used for *IR10* data, where multiple filing is more common.
[28] Opening book values are reported directly by the respondent on the AES form, but these are not included in the *LBD* dataset. This is the case for other variables also, as the data loaded into the *LBD* is the same as that provided by the *AES* team to the National Accounts team. *AES* processing creates variables useful for calculating GDP, etc. but also removes some raw line items.

4.1. Population and data-availability restrictions

Table 2 shows the impact of the industry and sector restrictions that we impose. The upper panel is restricted to *PENT*s that have positive employment in at least one month, and shows the loss of *PENT*s and loss of FTE employment due to the population restrictions. The numbers are unrounded averages, based on rounded annual counts for the period 2001 – 2012. The main reductions are due to the exclusion of non-market firms (11% of *PENT*s and 6% of employment), and the exclusion of firms that are not private-for-profit firms (6% of *PENT*s and 25% of employment). There is a very small loss of firms and employment due to missing information on industry or sector. The average number of included *PENT*s per year is 292,978.

The lower half of Table 2 shows the loss of sales that is due to the population restrictions, for firms that report positive sales. The measure of sales is based on *GST*-based *BAI* (Business Activity Indicator) data. The *BAI* data are available for a larger set of firms than those for which fuller production information is available. Excluding firms that never employ results in a loss of 21% of *PENT*s, and 7% of sales. There are also losses of 14% of *PENT*s and 17% of sales as a result of market and sector restrictions.

Table 2: Impact of population restrictions (Average per year: 2001-2012)

	Total	Included	Never L>0	No ind/sector	Non-market	Not Private-for-profit
			PENTs with positive employment			
Number of PENTs	353,766	292,978	0	1,211	39,547	20,029
		83%	0%	0%	11%	6%
FTE employment	1,401,697	970,433	0	63	78,942	352,258
		69%	0%	0%	6%	25%
			PENTs with positive sales			
Number of PENTs	495,590	318,763	106,279	415	48,500	21,633
		64%	21%	0%	10%	4%
Total sales ($m)	462,674	348,917	34,105	24	15,952	63,676
		75%	7%	0%	3%	14%

Note: Counts are unrounded averages based on rounded annual counts for 2001-2012. Underlying annual PENT counts are randomly rounded to base 3 prior to averaging. Underlying employment counts are rounded using graduated random rounding prior to averaging.

Requiring production information from *AES* or *IR10* sources reduces the number of useable observations. As shown in Table 3, an average of 94,404 *PENT*s per year have no production information. This amounts to 32% out of the average 292,978 with positive employment, as shown in Table 2. There is thus an average of 198,573 *PENT*s per year with production data, 96% of which have production data only from *IR10*, 2% with data from *AES* only, and 2% for which data are available from both sources. The absence of production data is largely due to the fact that there are alternative methods, other than filing an *IR10*, of satisfying the Inland Revenue filing requirements.

Table 3: Useable observations – by data source (Average per year: 2001-2012)

Pent-year observations	No production information (% of total)	Source of production information AES	IR10	Both	Total with production information
Pent with a single enterprise	94,119	3,088	189,559	4,676	197,323
	(32%)	2%	96%	2%	100%
Pent with multiple enterprises	285	11	1,217	21	1,248
	(19%)	1%	97%	2%	100%
Total	94,404	3,104	190,778	4,691	198,573
	(32%)	2%	96%	2%	100%

Note: Unrounded averages based on rounded annual counts. Underlying annual counts are randomly rounded to base 3 prior to averaging.

There is a relatively small number of *PENT*s that contain more than one enterprise in a year. For these *PENT*s, we choose the most consistent available data source. In cases where some enterprises have production data from only one source and others have information from both sources, we rely on the single data source, even though this means discarding *AES* data in some cases. We discard *PENT*s for which there is no common data source across all enterprises. *PENT*s may, however, have *AES*-sourced information in one year and *IR10*-sourced information in other years. Table 3 shows that the number of 'multiple enterprises' cases is relatively small.

4.2. Choosing a useable time span for the data

The derived dataset used for estimation covers the maximal period but does not use data for all periods for which some data are available in the *LBD*. The binding restrictions apply to the first and last periods of the data.

The initial period is limited by our ability to identify whether a *PENT* was operating in the previous year, in order to calculate consistently the value of capital assets (see below). We restrict the first year of productivity estimation to be the 2000/2001 financial year (which we will refer to as the 2001 year), which is the first year for which lagged data are available.

The final useable year of data will vary according to when the dataset is created. The binding constraints are generally the availability of *AES* and *IR10* data, and the ability to identify working proprietors based on annual tax returns. As an example, at the time of writing, the most recent complete archive of the LBD was from December 2014 (ibuldd_clean_archive_dec_2014). The most recent *AES* data relate to the 2012 year (dim_year_key=201203). Later data are available for some of the data sources (eg, *EMS* and *GST*) but this is insufficient for constructing the productivity dataset.

4.3. Capital

There are two main challenges in deriving a capital input measure. First, we want a measure that can be calculated consistently across the two financial data sources (*AES* and *IR10*).

Second, we want a measure that provides a consistent indication of capital use for firms that lease capital inputs as well as for those that own their capital inputs. Our general approach to measuring capital inputs is to estimate the flow of capital services used by the firm in a year.[29] We estimate three components of capital services flows:

$$\begin{aligned} &Value\ of\ capital\ services \\ &\qquad = depreciation + rental\ and\ leasing\ costs + cost\ of\ borrowing \end{aligned}$$

Both *AES* and *IR10* collect book value information for various classes of fixed assets. Asset categories differ across *AES* forms but can be aggregated to match *IR10* categories.

Depreciation

AES forms collect balance sheet information, including opening and closing book values and depreciation, for the various classes of fixed asset listed in Table 4. *IR10* forms collect a direct measure of depreciation costs during the year (*i10_deprec*), though not separately by asset type.

Table 4: Asset groupings in AES and IR10

Asset class	AES components Book values are bvf<type>_amt, where <type> is one of:	IR10 components
Vehicles	• Mv (motor vehicles) • Plane • Bus • Ships	I10_vehicles
Plant and Machinery (including other fixed assets[a])	• Pme (plant, machinery, equipment and other) • Lift • Hard (hardware) • Soft (software)	I10_plantmac +I10_othfass
Furniture and Fittings[b]	• Ffurn	I10_furnfitt
Land and Buildings	• Land • Li (Land improvements) • Nrb (non-residential building) • Rb (residential building) • Oc (Other construction)	I10_landbld

[a.] *'Other assets' are pooled with plant and machinery as part of AES processing*
[b.] *Furniture and fittings data are not collected for the Forestry and Logging industries (form FL) or Gas Supply (form GA) but some observations have non-zero data. This is transferred to Plant and Machinery. Affected industries are identified as those where nzsioc_lvl3='AA21' or (nzsioc_lvl3='AA32' and anz96_4d like 'A03%')).*

In *AES*, we combine information on the consumption of capital and the amortisation of intangible assets. The *AES* variable *fact_aes_enterprise_year.cons_cap_amt* is the sum of depreciation and amortisation across all classes of asset. We obtain a measure of depreciation of all tangible assets by subtracting from this sum the amortisation of intangible assets

[29] The permanent inventory method (*PIM*) is often used internationally. We have insufficient investment data to follow this method, because investment data are restricted to AES, and are available only from 2001. *PIM* also captures only owned capital, so does not adequately capture the use of leased capital.

(*fact_aes_enterprise_year.amort_amt).*[30] Depreciation costs may be based on rates defined for tax purposes, though in New Zealand, these depreciation rates are largely set to approximate economic depreciation rates (see Fabling, Gemmell, Kneller, & Sanderson, 2013).

Cost of borrowing

The cost of borrowing is estimated as a user cost of capital, multiplied by average capital stock. We use a constant value of 10% for the user cost of capital, which is the average business base lending rate.[31] This rate is applied to the average stock of fixed assets, as captured by opening and closing book values.

The IR10 form collects only closing book values, so the average capital stock is calculated by averaging closing values for the current and previous year. Where prior-year data are not available, the current-year closing value is used. AES collects both opening and closing values on the same questionnaire. Where consecutive years of AES data are available, we average consecutive closing values. Otherwise, we average the reported opening and closing values. For firms in their first year of operation, as identified by the absence of prior GST or employment activity, the value of the opening capital stock is assumed to be zero.

Rental, leasing and rates costs

We include capital rental and leasing costs in our measure of capital services to capture the flow of capital services used by firms that do not own their capital. We include council rates to improve consistency, as rates are generally included in lease payments but reported separately as an expense by owners.

Because rental and leasing costs are identified separately in IR10 forms but not in AES forms, we impute AES rental and leasing costs using IR10 information. Imputation is based on the share of rental and leasing costs in expenditure. One difficulty in implementing this approach is that total expenditure is not captured consistently across the two data sources. Our pragmatic response is to use a subset of expenditures that can be measured consistently – both as the denominator for the dependent variable, and as the expenditure covariate. The choice is necessary for estimation purposes, though the subtotal has no particular economic or accounting basis.[32]

The IR10 measure of purchases includes purchases for resale, so we combine the separate AES variables for purchases and purchases for resale. We want an *AES* measure that includes rental and leasing costs but these are measured as part of 'all other expenditure', which is combined with purchases in the variable *purch_amt*. We consequently need to aggregate a range of *IR10* expenditure items that do not have a matching line-item in the *AES* data. Finally, we want a measure that includes road user charges because these are not separately identified in *IR10* forms. We must therefore combine all *AES* components that are included in the *IR10* measure of rates and other expenses (see section 4.4.1).

[30] In cases where this difference is negative, we calculate total depreciation on tangible assets by summing depreciation for each reported asset class.
[31] http://www.rbnz.govt.nz/statistics/tables/b3/. The 'business base lending rate' has been renamed by the Reserve Bank as the 'SME overdraft rate'.
[32] Appendix Table 3 summarises the calculation of the commonly measured expenditure.

The imputation of rental, leasing and rates costs (*RLR*) is based on the ratio of *RLR* costs to commonly measured expenditure (*CME*), as recorded in *IR10*s:

$$\lambda_{IR10}^{RLR} = \frac{RLR_{IR10}}{CME_{IR10}} \quad (9)$$

For firms that have both *AES* and *IR10* data in the same year, we use the firm's own reported ratio to impute for *RLR* in *AES* $\left(\widehat{RLR}_{AES} = \lambda_{IR10}^{RLR} * CME_{AES}\right)$. For firms that have *IR10* data available in only some years, we use a firm-specific average ratio, applied to CME_{AES} in years when only AES data are available. We use the same approach to impute $\widehat{RLR}_{I10}$ for zero or missing rental and leasing costs in *IR10* – using firm averages where available. For the remaining firms that never report positive RLR_{IR10}, we impute using a regression model.

We run industry-specific regressions to estimate the share of *CME* accounted for by *RLR*, as a function of variables that can be measured in both the *AES* and *IR10* data. The common covariates that we use are depreciation, capital book values, expenditure, interactions of book values and expenditure, and year effects.[33]

We estimate the regression using *IR10* data, and use the coefficients to predict rental and leasing costs for *AES* data, and for *IR10* records where rental, leasing and rates information is zero, and therefore potentially missing.[34] Where the predicted *RLR* amount exceeds the sum of the components where these expenses would be reported, we replace the prediction with the sum of the components. For *AES*, these components are 'other indirect taxes' and 'other expenses'. For *IR10*, they are 'vehicle expenses' and 'other expenses'.

The estimated *RLR* amount is transferred from intermediate consumption to capital services.

4.3.1. Price deflation

In order to convert the expenditure-based capital services measure into an index of the quantity of capital services, we deflate at the aggregate level using the Capital Goods Price Index. Using an aggregate index does not control for cross-industry variation in the price of capital goods used in different industries but it is the only capital price index available.

4.4. Output

In the AES data, we use the measure of gross output contained in the variable *fact_aes_enterprise_year.gros_out_amt*.

When using *IR10* data, gross output is initially approximated by total income, adjusted for change in stocks and excluding income from interest and dividends:

$$Gross\ output = sales + other\ income + stock\ change = i10\_salesserv + (i10\_rent\_rcvd + i10\_otherinc) + (i10\_clgstock - i10\_opgstock) \quad (10)$$

[33] The regression is estimated using Stata's *glm* command, with parameters '*link(logit) family(binomial)*'.

[34] Rental leasing and rates costs should be greater than zero, so reporting zero in *IR10*s suggests an incorrect itemisation in the expense return.

4.4.1. Adjustments to IR10

We then adjust the *IR10* gross output measure to make it more directly comparable with the *AES* benchmark measure. Specifically, we adjust for purchases of goods for resale, for interest payments in financial industries, and for road user charges.

Interest payments

For general finance and insurance industries, *AES* treats net interest received as part of gross output, whereas in the *IR10* form, interest received and interest paid are recorded as components of income and expenses respectively. *IR10* records are adjusted so that the treatment is the same as in AES, adding net interest payments to gross output. This adjustment affects industries that receive General Insurance (*GI*) and Financial Services (*FS*) versions of the AES forms. In the past, these were allocated on the basis of ANZSIC96 industries (F4621, K73, K7422, K75). We apply the adjustment based on the corresponding ANZSIC06 codes (K62, K64, K6322, and F3501).[35]

Purchases of goods for resale

The *AES* measure of gross output excludes purchases of goods for resale from gross sales, in accordance with standard national accounts conventions. An examination of industry-by-industry differences in reported sales for firms with both *AES* and *IR10* records suggests that in some industries, many firms report resale purchases in *IR10*s as part of intermediate consumption.

Using *AES*, we calculate, for each productivity industry (*PF_IND* in Appendix Table 1) and year, the share of resale purchases (*presale_amt*) in the sum of *CME*, net of estimated *RLR* as described above. We then impute resale purchases in the *IR10* data by applying that share to the equivalent expenditure measure:

$$\lambda_{AES}^{resale} = \left(\frac{\text{total resale purchases}_{AES}}{\text{CME}_{\text{AES}} - \widehat{RLR}_{AES}}\right) \tag{11}$$

$$\text{Estimated resale purchases}_{IR10} = \lambda_{AES}^{resale} * \left(\text{CME}_{\text{IR10}} - \widehat{RLR}_{IR10}\right)$$

Estimated resale purchases for *IR10* observations are then subtracted from both gross output and from intermediate consumption.

Road user charges

Road user charges should be excluded from intermediate consumption but are not separately identified in *IR10* forms. We therefore remove a proportion of *IR10* intermediate consumption. The proportion is calculated as the proportion of *CME* in *AES*, net of estimated *RLR* that is accounted for by (separately reported) road user charges. This ratio is calculated by productivity industry and year, and applied after adjusting for interest payments and resale purchases.

[35] The interest adjustment is also relevant for firms in parts of the Property Services industries (ANZSIC06=L67; ANZSIC96=L7719, L7729, L773). However, this industry is excluded from our analysis because the associated AES form does not collect balance sheet information.

This adjustment is most relevant for transport-related industries.

$$\lambda_{AES}^{RUC} = \left(\frac{\text{road user charges}_{AES}}{\text{CME}_{\text{AES}} - \widehat{RLR}_{AES}}\right) \tag{12}$$

$$\text{Estimated road user charges}_{IR10} = \lambda_{AES}^{RUC} * \left(\text{CME}_{\text{IR10}} - \widehat{RLR}_{IR10}\right)$$

Estimated road user charges for *IR10* observations are then subtracted from intermediate consumption.

4.4.2. Price deflation

In order to convert the revenue-based output measure into an index of output quantity, we deflate at the level-3 *NZSIOC* industry level using the PPI for outputs. Where output prices vary due to varying levels of competition or markup pricing, or selling in different markets, this industry-level control may be inadequate (Klette & Griliches, 1996; Martin, 2008).

4.5. Labour

Until 2015, employment information in the *LBD* came from information extracted from monthly *LEED* data, with an estimate of the number of employees working at each firm on the 15th of each month, based on imputed job start and end dates. Derived *LEED* data are no longer available to researchers and an alternative method of measuring labour input has been developed making use of raw *IR EMS* data. The method is documented in more detail in Fabling & Maré (2015). Total labour input is calculated as the sum of employee labour input and working proprietor labour input.

The primary source for employee data is table *ir_clean.ird_ems* from Statistics New Zealand's *IDI*, which contains monthly job information for each worker in each *PBN* (plant) for which *PAYE* deductions are made. The dataset used in this paper makes use of the December 2014 instance of the *IDI*. From this information, we identify the number of employees in each *PENT* in each month, distinguishing between 'terminal months' (start or end months within a job) and 'interior months'. Interior months are adjusted to reflect *FTE* employment (described below), and terminal months are then adjusted based on the adjacent internal months. This is done to allow more accurate treatment of part-month employment.

Working proprietors are identified from annual tax return information in the *IDI*. An individual is identified as a working proprietor if they receive self-employed income.[36] If an individual is *ever* a shareholder with a *PENT*, they are treated as a shareholder in all periods in which they are employed, and do not count towards the employee labour input. Instead, their apparent labour input from the *PAYE* information is counted as working proprietor labour input. If an individual is a working proprietor in more than one *PENT*, they are assumed to provide an equal fraction of their labour input to each *PENT*.

[36] Self-employed income is identified as: non-zero net profit reported in an *IR3*; *IR7P* with non-zero total income and *IR3* partnership income; *IR4S* non-zero remuneration. We disregard shareholder salary income below a minimum threshold (the threshold is set at $15,000 in 2000, and adjusted for CPI movement). The threshold aims to exclude non-working shareholders and non-owner directors.

4.5.1. Full time equivalent employment

Our preferred measure of labour input adjusts the number of employees to approximate a *FTE* measure. This not only provides a more appropriate measure of labour input but also allows for monthly earnings to be adjusted to more closely approximate a monthly wage rate. The adjustment entails the following steps:[37]

a) If monthly earnings in a job are lower than what would be earned working full time at the relevant (adult or youth) minimum wage, *FTE* is calculated as the ratio of monthly earnings to full-time minimum wage earnings;
b) If an employee has multiple jobs in a month, the sum of *FTE* is constrained to have a maximum of 1, by proportionally scaling *FTE* in all jobs in that month based on earnings in each job.
c) Total monthly *FTE* is reduced if the employee receives income from benefits that are generally associated with lower employment intensity (identified as ir_ems_income_source_code in (*'BEN'*,*'CLM'*,*'PPL'*,*'STU'*), which relate to working age benefits, *ACC* payments, paid parental leave, and student allowances)
d) *FTE* is adjusted in employees' first and last months in a job – using average implied wage rates from the two adjacent interior months[38].
   i) If first or last month earnings are less than the implied wage rate, earnings in the first or last month is divided by the wage rate in adjacent months. E.g.: for the starting month (month 1), $\widehat{FTE_1} = \frac{earnings_1}{(earnings_2/FTE_2)}$.
   ii) Actual reported end dates are used if this suggests a lower *FTE*.
   iii) Terminal months associated with 'short spells', where information from adjacent months is not available, and unadjusted end dates are assumed to have FTE of at most one half.

Monthly labour input is aggregated to *PENT* level and aligned to the *PENT*'s balance date-year, to ensure that labour input is aligned with firm financial measures.

4.6. Intermediate consumption

Intermediate consumption in AES is measured by the variable *fact_aes_enterprise_year.ic_adj_amt*. With IR10 data, we construct an analogous measure as the sum of purchases and total expenses, excluding salaries and wages, bad debt write-offs, interest paid, and depreciation:

$$ic\_i10 = i10\_purchases + i10\_totexp\ sales - i10\_salwages - i10\_baddebts - i10\_intpd - i10\_deprec$$

[37] This FTE adjustment differs from that in Maré & Hyslop (2006), by using within-spell information to adjust start and end months, and by more accurately identifying working proprietor input.
[38] Jobs are assumed to continue if there is a one month gap in the data. We use a two month window for calculating implied wage rates to increase the number of start and end months for which interior-month data are available.

For both *AES* and *IR10*, intermediate consumption is adjusted, as described above, by the following adjustments:

$$ic\_i10^* = ic\_i10 - goods\ for\ resale - road\ user\ charges$$
$$- (rental\ leasing\ and\ rates\ if\ RLR\ is\ missing)$$

$$ic\_aes^* = ic\_aes - (estimated\ rental\ and\ leasing\ )$$

For AES only, a proportion of the RLR amount is also subtracted from intermediate consumption, because rates are separately reported and are not included as part of intermediate consumption. The proportion is estimated from IR10 responses as the ratio of rates expenditure to the sum of rates and rental and leasing costs.

$$\lambda_{I10}^{Rates} = \frac{Rates_{I10}}{RLR_{I10}} \tag{13}$$

This is applied to AES data at the firm-year level for dual filers, using a firm-average where available, or else using an industry-year average ratio.

$$IC_{AES} = IC_{AES} - (1 - \lambda_{I10}^{Rates}) * \widehat{RLR}_{AES} \tag{14}$$

4.6.1. Price deflation

In order to convert the expenditure-based intermediate consumption measure into an index of the quantity of intermediate inputs, we deflate at the level-3 NZSIOC industry using the PPI for Inputs.[39] Where firms face heterogeneous input prices or qualities, due to buying in different locations or from imperfectly competitive suppliers, this industry-level control may be inadequate (see, for instance, Grieco et al., 2013).

5. Using the data for statistical analysis

The methods described so far aim to generate a robust and consistent dataset of production data for as large a set of *PENT*s as possible. Inevitably, some anomalies remain. These may be due to reporting or recording errors, to imperfections in our data manipulation and linking, or to anomalous real-world changes for particular firms. In this section, we describe the removal of outliers, and note the merits of testing the sensitivity of statistical analyses on the basis of data sources or industry to ensure that findings are not unduly influenced by the way that the data have been assembled.

[39] For the calculation of value added, which is the difference between gross output and intermediate consumption (see Figure 1), gross output and intermediate consumption are thus deflated using different deflators.

5.1. Data issues

Removing outliers and anomalies

Before using the constructed dataset for statistical estimation, it is advisable to remove some obvious outliers. We do not want to discard legitimate data variation but neither do we want to include data that is patently incorrect. Our approach is to apply pragmatic rules to remove potentially questionable data. We have performed spot-checks on the identified outliers and concluded that some are due to clear errors in reporting, are consequences of data processing, or reflect atypical events at the plant or enterprise level.

Our estimation sample excludes all years of data for firms that experience a large annual change in either inputs or outputs. We identify large changes on the basis of numeric and proportional (log) changes. Specifically, we calculate annual log changes for gross output, capital services, intermediate consumption and employment, and identify cases where the absolute value of any of these changes exceeds four. If the associated numeric change is also large (an absolute change of 20 in employment, or of $50,000 in other variables), we discard all observations for the affected *PENT*s. We also exclude observations where our adjustments to output, intermediate consumption, or capital services result in zero, negative, or missing values for these variables.

The combined impact of these exclusions is to reduce the average number of *PENT*s per year by about 4%, from 198,573, as shown in Table 3, to 190,642 (a rounded total of 2,287,707 *PENT*-year observations).[40]

Distinguishing data sources

Researchers may wish to check for the sensitivity of their estimates to the source of data used and the method of adjustment used. To facilitate such analyses, the *datasource* variable is generated, which allows the researcher to exclude, or analyse separately, observations that use modelled or imputed data. This is a categorical variable that identifies the primary source of data, and whether or not rental and leasing costs have been imputed. Table 5 shows the different values of the variable, and the proportion of observations and employment accounted for by each combination of data source and adjustment.

It is not uncommon for *PENT*s to have different sources of data or different adjustments applied in different years. Around two-thirds of *PENT*s that ever have annual data based primarily on *AES* data also have at least one year where their data are based on *IR10* data only. This has the potential to introduce within-firm variation in *mfp* estimates. Where these differences are substantial, the *PENT* will be dropped as an outlier. The *datasource* variable allows researchers to examine whether estimated changes in *mfp* are associated with changes in data or adjustments.

[40] All annual counts of *PENT*s are randomly rounded to base 3.

Table 5: The 'datasource' indicator

Datasource	AES	IR10	RLR imputed	missing IR10 lag	% of PENT-Year obs	% of FTE employment
Aes_avg	In_aes==1	In_i10==0	Rlr_imp=avg		1%	13%
Aes_mod	In_aes==1	In_i10==0	Rlr_imp=model		1%	35%
Both_own	In_aes==1	In_i10==1	Rlr_imp=own		0%	1%
Both_avg	In_aes==1	In_i10==1	Rlr_imp=avg		0%	0%
Both_mod	In_aes==1	In_i10==1	Rlr_imp=model		2%	13%
I10_own_nolag	In_aes==0	In_i10==1	Rlr_imp=own	Nolag_i10==1	11%	2%
I10_avg_nolag	In_aes==0	In_i10==1	Rlr_imp=avg	Nolag_i10==1	2%	0%
I10_mod_nolag	In_aes==0	In_i10==1	Rlr_imp=model	Nolag_i10==1	12%	1%
I10_own_lag	In_aes==0	In_i10==1	Rlr_imp=own	Nolag_i10==0	2%	0%
I10_avg_lag	In_aes==0	In_i10==1	Rlr_imp=avg	Nolag_i10==0	61%	31%
I10_mod_lag	In_aes==0	In_i10==1	Rlr_imp=model	Nolag_i10==0	9%	5%

Working-proprietor-only firms

The dataset contains enterprises that have no employees. The inclusion of separate employee and working proprietor counts allows researchers to test for differing production parameters for employing and working-proprietor-only firms, or to derive estimates for employing firms only.

Industry disaggregation

Different firms operate with different technologies. With panel data, it is possible to estimate firm-specific production function parameters – at least for firms that have a sufficient number of annual observations. It is customary, however, to estimate common parameters for firms using similar technologies – usually by industry. Productivity is then estimated relative to the mean productivity within the industry.

There is no 'best' level of industry aggregation at which to estimate production function parameters. Choosing a level of aggregation entails weighing up precision (from larger samples) and bias (from pooling firms with dissimilar technologies). Parameter estimates for very small groups of firms are likely to be imprecisely estimated, and also risk being disclosive.

PPI indexes are available separately for ANZSIC06 or NZSIOC industry groupings, making this an obvious basis for categorising firms. The AES survey is stratified by ANZSIC06 industries (ANZSIC96 prior to 2006). Deriving separate estimates on the basis of strata is also sensible, and has the added advantage of limiting the extent to which data collected using different AES forms is combined.

When estimating separate production parameters for different NZSIOC groupings, we group some level-3 NZSIOC06 categories to avoid estimation based on small samples. This is advisable for statistical reasons, as well as avoiding disclosure risks. This is, however, done after deflating by the PPI at the more detailed level. The following NZSIOC industries are grouped at level 2 rather than level 3: CC1; CC3; CC5; CC7; DD1. In addition, two other level 2 industries are grouped together (KK11 and KK12) and are labelled as 'KK1_'. Our imputations and adjustments are applied at this productivity-industry (*pf_ind*) level.

Potential limitations

The final productivity dataset has been developed with the purpose of supporting productivity research. It is worth, however, noting some potential limitations of the data:

- Official statistics: The dataset is not necessarily representative of the population of firms (see below). It can support analysis of variation across firms, and of relationships between variables at the firm level but caution is needed in generalising findings.
- Subsets of firms: Caution should be exercised when analysing patterns for small subsets of firms – for both statistical and disclosure reasons. At various stages, imputation and deflation are applied at the industry level – generally level-3 NZSIOC. When analysing data from firms that belong to different industries, allowance should be made for these industry-level adjustments (eg: industry-specific coefficients or error components).
- Change and dispersion: There is a good deal of variability in production data – both cross-sectionally, and for individual firms over time. It is difficult to distinguish genuine variability from that generated by data issues. Inconsistent reporting – across firms, over time, and between data sources will affect measured variation, as will our data processing procedures to some extent. Research that focuses on variability, such as studies of productivity dispersion or firm-level productivity growth may be particularly affected by non-sampling variability.

5.2. Codebook

Table 6 contains a list of variables that are included in the final productivity dataset *pent_prod_IDI_20141205*, which is stored within the *ibuldd_research_datalab* database on the *IBULDD_clean* server [*wprdsql31\ibuldd*].

Table 6: List of variables in pent_prod_IDI_20141205

Number of rows (rounded): 2,287,707
Number of variables: 19

Variable name	Variable type	Description	values?
pent	char(10)	Permanent enterprise number	Code
dim_year_key	int	Year identifier	YYYY03
active_prior_year	tinyint	Indicator for whether the pent was active in the previous year	0/1
active_next_year	tinyint	Indicator for whether the pent is active in the subsequent year	0/1
datasource	varchar(13)	Identifies the source of the financial data (See section 5.1)	Code
nzsioc_lvl3	char(4)	Level 3 NZ Standard Industry Output categories (NZSIOC06)	Code
pf_ind	varchar(4)	Grouped productivity-industry identifier, based on NZSIOC.	Code
anz96_4d	char(5)	4-digit ANZSIC96 industry code	Code
anz06_4d	char(5)	4-digit ANZSIC06 industry code	Code
fte	float	Labour input from employees – full-time equivalents.	Magnitude
wp	float	Labour input from working proprietors	Magnitude
lngo	float	Natural log of gross output	Value
lnm	float	Natural log of intermediate inputs	Value
lnk	float	Natural log of capital services	Value
lnl	float	Natural log of labour input (fte+wp)	Value
mfp_go_cd	float	Estimate of multi-factor productivity from a *pf_ind*-specific gross output Cobb-Douglas production function	Value
mfp_go_tl	float	Estimate of multi-factor productivity from a *pf_ind*-specific gross output translog production function	Value
mfp_go_fe	float	Estimate of multi-factor productivity from a *pf_ind*-specific gross output Cobb-Douglas production function with *PENT* fixed effects (fixed effects are included in the *mfp* measure)	Value
go_fe	float	Estimate of the firm fixed effect from a *pf_ind*-specific gross output Cobb-Douglas production function with *PENT* fixed effects	Value

Note: Table is stored with primary key clustered (pent,dim_year_key)

Table 7 shows the average number of observations per year, by the industry grouping used for data adjustments and for the calculation of the industry-specific *mfp* measures included in the dataset. The industries are sufficiently aggregated to meet confidentiality requirements when data are pooled across years. However, data for some industries (AA32 and JJ12) fail confidentiality checks in some years, so disaggregated analyses should always be carefully checked.

Table 7: Average annual observation count and FTE total [pent_prod_IDI_20141205]

pf_ind	*PENT*s	FTE
AA11: Horticulture and fruit growing	6,045	8,940
AA12: Sheep, beef cattle, and grain farming	18,570	5,973
AA13: Dairy cattle farming	10,020	8,568
AA14: Poultry, deer, and other pstock farming	3,603	1,908
AA21: Forestry and logging	1,299	3,051
AA31: Fishing and aquaculture	957	723
AA32: Agric, forest, fish support services, and hunting	3,978	9,243
BB11: Mining	225	3,483
CC1: Food, Beverage, Tobacco Mfrg	1,881	60,156
CC2: Textile, leather, cloth, and footwear manufacturing	1,398	11,001
CC3: Wood and Paper product manufacturing	1,635	18,825
CC41: Printing	966	7,173
CC5: Chemical, rubber, non-metallic mfrg	774	17,775
CC61: Non-metallic mineral product manufacturing	528	5,973
CC7: Metal and metal product manufacturing	2,376	18,900
CC81: Transport equipment manufacturing	993	6,927
CC82: Machinery and other equipment manufacturing	2,412	20,541
CC91: Furniture and other manufacturing	1,962	7,257
DD1: Electricity gas supply & water	552	8,991
EE11: Building construction	9,621	15,159
EE12: Heavy and civil engineering construction	912	19,233
EE13: Construction services	19,278	33,402
FF11: Wholesale trade	9,204	60,342
GH11: Motor vehicle & parts, and fuel retailing	2,118	16,569
GH12: Supermarket, grocery, and specialised food retailing	3,549	32,535
GH13: Other store-based and non-store retailing	11,310	58,719
GH21: Accommodation and food services	11,079	48,642
II11: Road transport	5,607	19,926
II12: Rail, water, air, and other transport	642	15,402
II13: Post, courier support, and warehouse services	2,724	22,185
JJ11: Information media services	1,542	11,208
JJ12: Telecommunication, Internet, and library services	336	11,325
KK13: Auxiliary finance and insurance services	2,280	7,875
KK1_: Finance Insurance and superannuation	876	33,789
LL11: Rental and hiring services	2,070	6,591
MN11: Professional, scientific, and tech services	25,461	60,693
MN21: Administrative and support services	7,920	39,507
RS11: Arts and recreation services	3,210	8,367
RS21: Other services	10,728	22,932
TOTAL	190,641	769,809

Note: Unrounded average based on rounded annual counts. Numbers are averages for the years 2001-2012.

5.3. Comparison with official statistics

The dataset of useable data is incomplete, and should not be expected to match aggregate statistics. Although the concepts and definitions are not exactly comparable,

Table 8 compares aggregate total income and employment from our productivity dataset with published aggregates for a single year – the year to March 2012. Within the industries that are covered in the productivity dataset, aggregate total income from our data is 62% as large as officially measured total income (*AES)*. This varies greatly across industries, ranging from 29% to 96%.

Table 8a: Total income for the year to March 2012 - coverage by industry

	pf_ind grouping	Official AES measure ($m) Total Income ($m)	Firms in productivity dataset (unweighted AES/IR10, $m) Gross Output ($m)	Coverage (prod/ official)
AA11	Horticulture and fruit growing	$3,196	$1,476	46%
AA12	Sheep, beef cattle, and grain farming	$9,508	$3,918	41%
AA13	Dairy cattle farming	$12,430	$6,098	49%
AA14	Poultry, deer, and other livestock farming	$1,560	$619	40%
AA21	Forestry and logging	$3,779	$1,386	37%
AA31	Fishing and aquaculture	$1,073	$362	34%
AA32	Agric, forest, fish support services, and hunting	$2,805	$1,973	70%
BB11	Mining	$7,844	$4,843	62%
CC1	Food and beverage manufacturing	$42,408	$39,004	92%
CC21	Textile, leather, cloth, footwear manufacturing	$2,480	$1,902	77%
CC3	Wood and paper product manufacturing	$7,841	$5,611	72%
CC41	Printing	$1,674	$1,131	68%
CC5	Petrochemical product manufacturing	$21,206	$17,512	83%
CC61	Non-metallic mineral product manufacturing	$2,535	$1,496	59%
CC7	Metal and metal product manufacturing	$9,039	$5,489	61%
CC81	Transport equipment manufacturing	$2,728	$1,668	61%
CC82	Machinery and other equipment manufacturing	$7,043	$5,713	81%
CC91	Furniture and other manufacturing	$1,596	$1,055	66%
DD1	Electricity, gas and water	$17,706	$15,217	86%
EE11	Building construction	$11,892	$6,907	58%
EE12	Heavy and civil engineering construction	$8,419	$6,035	72%
EE13	Construction services	$13,668	$6,834	50%
FF11	Wholesale trade	$81,925	$46,914	57%
GH11	Motor vehicle & parts, and fuel retailing	$13,402	$6,884	51%
GH12	Supermarket, grocery, and spec. food retailing	$18,887	$12,077	64%
GH13	Other store-based and non-store retailing	$24,129	$15,491	64%
GH21	Accommodation and food services	$10,322	$5,734	56%
II11	Road transport	$7,062	$5,062	72%
II12	Rail, water, air, and other transport	$7,252	$6,979	96%
II13	Post, courier support, and warehouse services	$6,326	$5,349	85%
JJ11	Information media services	$5,379	$3,650	68%
JJ12	Telecomm., Internet, and library services	$9,176	$7,674	84%
KK1_	Finance, insurance, and real estate	$76,547	$34,825	45%
KK13	Auxiliary finance and insurance services	$4,572	$2,512	55%
LL11	Rental and hiring services	$4,444	$2,161	49%
MN11	Professional, scientific, and technical services	$29,054	$18,135	62%
MN21	Administrative and support services	$8,890	$5,097	57%
RS11	Arts and recreation services	$6,888	$2,009	29%
RS21	Other services	$8,012	$3,127	39%
	TOTAL (in-scope industries)	$514,697	$319,928	62%

In

Table 8b, we compare our *FTE* employment measure with the measure published from the Quarterly Employment Survey (*QES*). The *QES* measure (sum of full-time employees plus half part-time employees) differs from our measure, and the industry coverage differs from that of the productivity dataset. For a consistent subset of industries covered in both datasets, our FTE measure implies employment that is 76% as large as that shown by the *QES*. For one industry, our measure is larger than the corresponding *QES* measure.

Table 8b: Production Aggregates for the year to March 2012: FTE employment

QES industry grouping	QES FTE employment	Derived FTE employment	Ratio (Derived/ Official)
AA: Agriculture	n/a	39,710	n/a
AB: Forestry and Mining	10,800	4,600	43%
AC: Manufacturing	171,500	155,200	90%
AD: Electricity, Gas, Water and Waste Services	10,900	12,300	113%
AE: Construction	90,300	72,100	80%
AF: Wholesale Trade	84,900	61,100	72%
AG: Retail Trade	132,400	112,400	85%
AH: Accommodation and Food Services	69,300	51,400	74%
AI: Transport, Postal and Warehousing	68,700	58,800	86%
AJ: Information Media and Telecommunications	28,700	23,200	81%
AK: Financial and Insurance Services	44,300	43,000	97%
AL: Rental, Hiring and Real Estate Services	21,300	6,900	32%
AN: Professional, Scientific, Technical, Administrative and Support Services	178,800	117,500	66%
AO: Public Administration and Safety	84,300	n/a	n/a
AP: Education and Training	120,700	n/a	n/a
AQ: Health Care and Social Assistance	145,500	n/a	n/a
AS: Arts, Recreation and Other Services	69,400	29,500	43%
AZ: Total All Industries	1,331,700	787,710	
Consistent subset of industries	981,300	748,000	76%

Notes: Official Employment measure is QES FTE employment (annual average for the year to March 2012). Official Output measure is from AES (year to March 2012).

6. Production function and productivity estimation for a selected industry: EE11 'Building Construction'

As noted in section 2, *mfp* is a measure of the efficiency with which firms convert inputs into outputs. More productive firms are those that produce more outputs for a given set of inputs. Equivalently, more productive firms are those that produce at a lower cost per unit of output.

We focus in this section on *mfp*, which takes into account all measured inputs into production. We do not discuss partial productivity measures, such as labour productivity (value added per worker; gross output per worker; revenue per worker).

Because *mfp* is inherently a relative measure, defined for a particular underlying technology, it is not possible to compare this measure of productivity across different technologies. A standard normalisation is that *mfp* has zero mean within each industry. To compare productivity across industries, we either need to impose constraints on differences in technology (for instance unrealistically imposing common production function parameters across industries), or make the comparison using partial productivity measures such as labour productivity (output per worker), which has a natural common metric.

In this section, we present estimates for one specific industry, to illustrate the use of the *LBD* data. The industry we have chosen is the 'Building construction' industry group (NZSIOC EE11). We have chosen this industry because it has a relatively large number of employing firms.

6.1. Examining differences across data sources

In this section, we document whether production function and *mfp* estimates depend on the primary source of data. Specifically, we examine the effect of including both *AES*-sourced observations and *IR10*-sourced observations. The comparison of estimates by data source provides insights not only into the influence of data sources, but also into the adequacy of the chosen production function and estimation method. Sub-group differences may arise for a range of reasons:

- Wrong functional form: The firms that are surveyed in *AES* are, on average, larger than those for which *IR10* data are used. Even if small and large firms operate with the same technology, estimated coefficients may differ by size, and hence by data source, if the fitted production function fails to capture non-linearities. If this is the case, more flexible functional forms should fit both subsamples better.
- Firm heterogeneity: Firms in different sub-samples may differ in productivity for unobserved reasons. If this is the case, estimation methods that allow for unobserved firm-level heterogeneity should reduce the apparent differences by data source.
- Data inconsistencies: Although every effort has been made to harmonise the variable definitions across data sources, it is possible that some systematic differences remain.

When a parsimonious model is estimated, there is a sizeable difference in mean *mfp* across the two samples. The first two columns of Table 9 show OLS estimates of the most parsimonious (Cobb-Douglas) production function for *AES* and non-*AES* observations respectively. The upper panel contains estimates from a value added production function. For that specification, the capital coefficient for *AES* firms is significantly larger than that of non-*AES* firms, and the explanatory power of the regression is lower for the non-*AES* sample, reflecting greater *mfp* variation within the larger pool of more diverse non-*AES* firms. The two samples are pooled in the third column of Table 9. Not surprisingly, the estimated coefficients are close to those of the much larger non-*AES* sample. We include an intercept to capture the mean difference between the two samples. Controlling for factor inputs, the *mfp* of *AES* firms is 0.444 (55%) higher than that of non-*AES* firms.

The nature of the difference is illustrated in the top left graph of Figure 2. That graph plots regression residuals (*mfp*) against fitted values from a pooled regression that does not include an *AES* dummy. Two smoothed lines are included to show mean *mfp* for the different samples. The line for *AES* observations is above the line for non-*AES* observations across the entire range of fitted values, with evidence of non-linearity in both lines.

The mean difference between *AES* and non-*AES* firms is reduced when the value added Cobb Douglas production function is estimated with firm fixed effects. The estimate in column 6 of Table 9 shows an estimated difference of 0.216. This reduction is consistent with there being time-invariant unobserved differences in productivity between *AES* firms and non-AES firms. When the unobserved differences are modelled as time varying, using the structural approach of Ackerberg, Caves & Frazer (ACF),[41] the estimated difference between datasets reduces further to 0.161 (column 9). The residual plot for these two additional specifications (not shown) looks very similar to the OLS plot shown in Figure 2, albeit with the two smoothed lines slightly closer together.

A more appreciable difference is evident when a gross output production function is used (rather than value added). The OLS residual plot for this specification is shown in Figure 2, below the value added plot. The smoothed lines are closer together, though still showing a pronounced curvature. The residuals are also much smaller on average than those from the value added specification, suggesting that treatment of intermediate consumption in the gross output specification is more appropriate.

The estimated coefficient on the *AES* indicator in a gross output production function is shown in the lower panel of Table 9. The OLS difference is reduced to 0.164 and the firm fixed-effect and ACF estimates are 0.087 and 0.052 respectively. Allowing for firm heterogeneity, by fixed effects or structural estimation, thus narrows the estimated difference between data sources but as is evident in the residual plots in Figure 2, there remains a distinct curvature, especially for *AES* observations. The regression underestimates output for smaller *AES* firms, resulting in relatively high estimated *mfp*.

41 We use a 3rd order polynomial of capital, labour, and intermediates in the control function. A 3rd order polynomial is chosen as the standard specification to accommodate the interaction terms in the translog specification. The coefficient on labour is not identified in the first stage regression. A 3rd order Markov polynomial is used in the second stage, using instruments created as powers of 3rd order polynomials in capital, lagged labour, and lagged intermediates. Estimation is done in two stages, with standard errors estimated from 50 bootstrap replications.

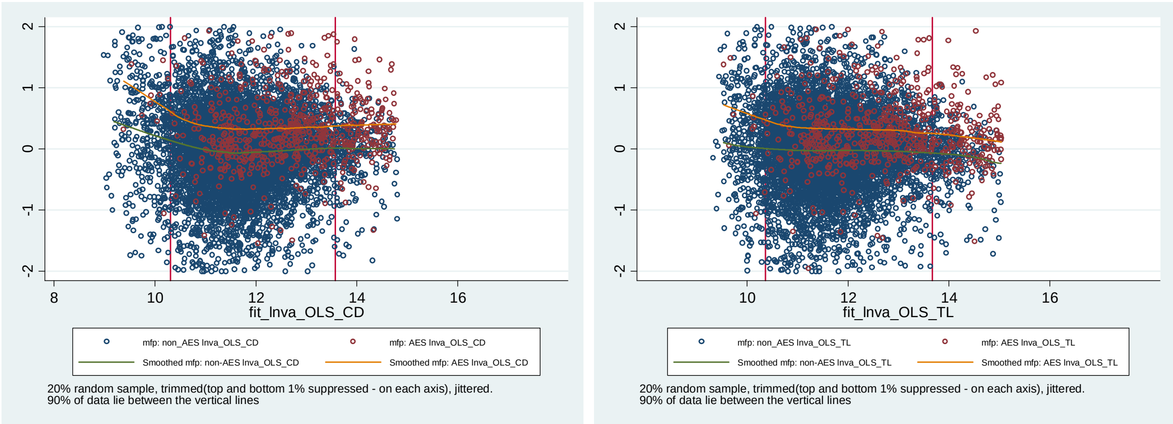
Table 9: Cobb-Douglas production function (industry EE11): Value added and Gross Output estimates

	OLS			Firm Fixed Effects			Ackerberg-Caves-Frazer		
	AES	Non-AES	All	AES	Non-AES	All	AES	Non-AES	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				(a) Value added production function					
ln(Labour)	0.564***	0.603***	0.603***	0.444***	0.541***	0.534***	0.669***	0.419***	0.412***
	[0.024]	[0.008]	[0.007]	[0.038]	[0.010]	[0.009]	[0.053]	[0.122]	[0.117]
ln(Capital)	0.398***	0.307***	0.316***	0.175***	0.252***	0.249***	0.233***	0.359***	0.405***
	[0.020]	[0.006]	[0.006]	[0.026]	[0.008]	[0.007]	[0.039]	[0.083]	[0.087]
AES			**0.444*** **			**0.216*** **			**0.161*** **
			[0.020]			**[0.017]**			**[0.011]**
Observations	3,768	46,011	49,779	3,768	46,011	49,779	2,934	30,600	33,534
R-squared	0.83	0.60	0.68	0.24	0.39	0.39	0.82	0.61	0.68
RTS	0.962**	0.910***	0.919***	0.618***	0.793***	0.784***	0.901***	0.778***	0.816***
	[0.018]	[0.006]	[0.006]	[0.038]	[0.009]	[0.009]	[0.023]	[0.046]	[0.039]
Number of firms				1,284	12,081	12,357			
				(b) Gross Output production function					
ln(Labour)	0.202***	0.210***	0.210***	0.182***	0.190***	0.190***	0.234***	0.273***	0.288***
	[0.022]	[0.005]	[0.005]	[0.022]	[0.005]	[0.005]	[0.024]	[0.051]	[0.025]
ln(Capital)	0.077***	0.052***	0.055***	0.025*	0.060***	0.058***	0.056	0.039	0.047*
	[0.014]	[0.004]	[0.004]	[0.014]	[0.005]	[0.004]	[0.039]	[0.025]	[0.025]
ln(Intermed)	0.691***	0.682***	0.683***	0.614***	0.670***	0.669***	0.690***	0.667***	0.646***
	[0.020]	[0.005]	[0.005]	[0.028]	[0.005]	[0.005]	[0.065]	[0.056]	[0.045]
AES			**0.164*** **			**0.087*** **			**0.052*** **
			[0.011]			**[0.007]**			**[0.003]**
Observations	3,828	47,139	50,970	3,828	47,139	50,970	2,964	31,197	34,164
R-squared	0.97	0.92	0.94	0.78	0.84	0.84	0.97	0.93	0.95
RTS	0.970***	0.945***	0.949***	0.821***	0.920***	0.917***	0.980	0.979	0.980**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.020]	[0.018]	[0.009]
Number of firms				1,305	12,288	12,567			

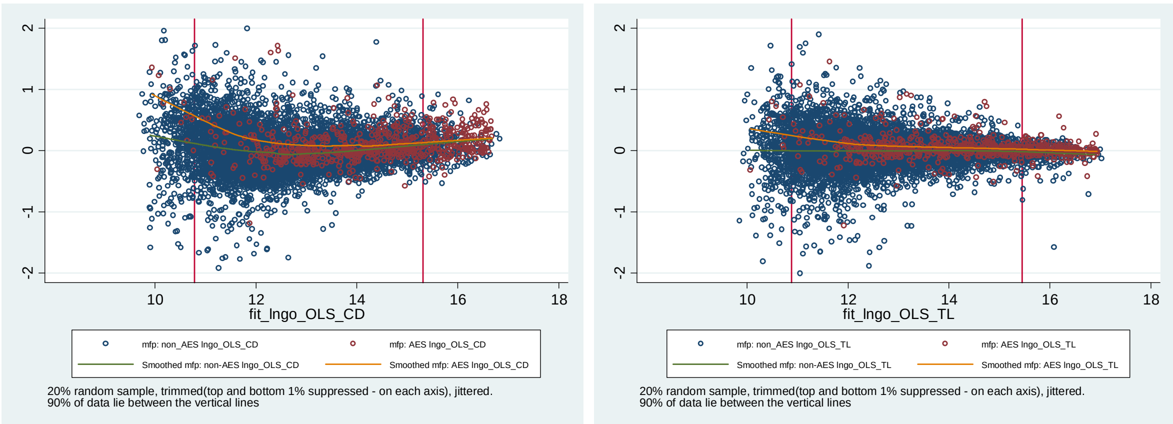
*Note: includes constant, year effects, *** p<0.01, ** p<0.05, * p<0.1. (1): RTS=Returns to scale. Significance indicators reflect difference from 1 (ie, constant returns to scale).*

Figure 2: *MFP (r*esidual) plots (industry EE11)

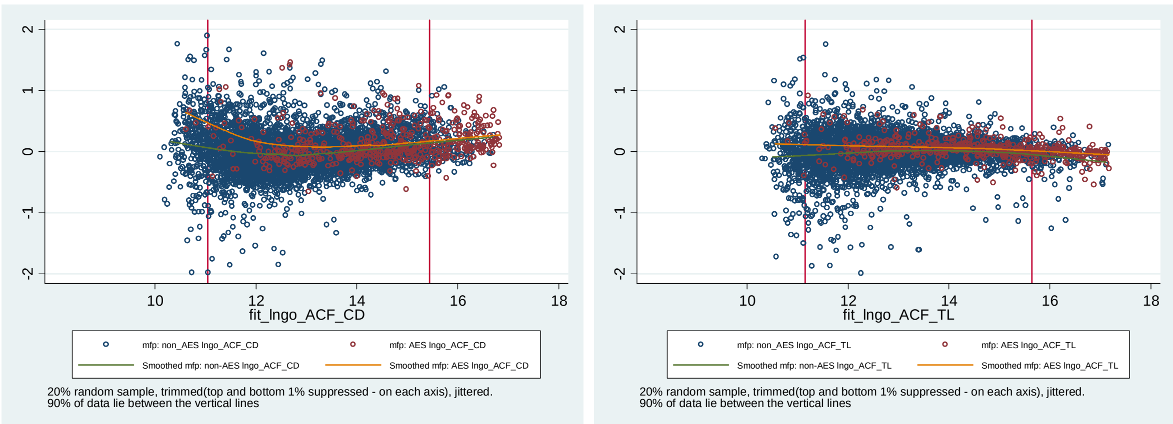
Value added production function: CD&TL (OLS)



Gross Output production function: CD&TL (OLS)



Gross Output production function: CD&TL (Ackerberg, Caves & Frazer estimation)



To examine whether this curvature and underestimation of output for small *AES* firms potentially reflect misspecification of the production function, we estimate the more flexible translog production function. The estimates are shown in Table 10. The first three columns show estimates from value added production functions and columns 4 to 6 show gross output production function estimates. With a value added production function, allowing for a more flexible functional form does little to reduce the estimated difference between *AES* and non-*AES* observations. The coefficients on the indicator for *AES* observations are similar to those estimated with the simpler Cobb-Douglas production function, showing a difference that ranges from 0.347 from the OLS estimation, to 0.161 for the ACF estimation. The top-right graph of Figure 2 shows the residual plot for the OLS specification, with a slightly smaller mean difference, and a slightly reduced curvature.

Allowing a more flexible (translog) functional form makes a more appreciable difference in the gross output specification, where intermediate consumption interacts with other inputs. As shown in the final three columns of Table 10, the coefficient on *AES* drops to between 0.052 (for *ACF*) and 0.058 (for *FE*). Furthermore, residuals from the gross output translog regressions not only have lower variance than the corresponding Cobb-Douglas estimates, but also display less curvature when plotted against fitted values. Residual plots are shown in the lower two graphs in the right-hand column of Figure 2. There is relatively little difference between the *OLS*, *FE* and *ACF* estimates (*FE* estimates not shown). All have relatively low variance and small curvature.

The residuals from the *ACF* gross output translog estimation appear to be the most well behaved – though the identification is not valid. Specifically, in the gross output translog case, the instruments used for identification cannot validly be excluded from the estimating equations.[42] This highlights one of the challenges of implementing structural estimation approaches for gross output production functions using flexible functional forms. The inclusion of intermediate consumption as a factor input means that it cannot act as an independent proxy variable. In principle, identification is still possible based on the higher order terms included in the polynomial term ($\omega(q,z)$ in equation 8). However, the translog production function itself includes second-order terms, so that identification must rely on the functional form of the polynomial, and the higher-order terms.[43]

Returning to our initial question of whether production function and *mfp* estimates depend on the primary source of data, we have shown that there are differences. The differences are most pronounced when we estimate them using a constrained production function specification (Cobb-Douglas, value added). The differences are much smaller when estimated from a flexible gross output production function. This highlights the general points made in section 2.2 that, conceptually, *mfp* reflects performance relative to other firms using the same inputs, with comparisons depending on what inputs are taken into account and how the relationship between inputs and outputs is modelled. A relationship between *mfp* and inputs, as is evident in the curvature of residual plots in Figure 2, indicates that the chosen production function is inadequate, in the sense that estimated *mfp* reflects performance differences between firms with different amounts of input, rather than between firms with the same inputs.

Of course, the danger in fitting more flexible functional forms is that, to the extent that unmeasured inputs are correlated with measured inputs, we run the risk of wrongly incorporating the influence of unmeasured inputs as a feature of a misspecified production function rather than as a component of *mfp* variation.

---

42 The ACF specifications in Table 10 report the value of Hansen's J statistic, and the associated p-value. The statistic tests for overidentification – whether the instruments are valid. A low p-value (below 0.05) indicates that the instruments cannot be excluded, and are thus *not* valid. The final row shows the number of bootstrap replications (out of 50). The value of 49 in the final column means that one replication failed to converge.

43 It is possible to find subsets and lags of the higher-order instruments that pass overidentification tests. The misspecified model is included here to highlight the practical challenges of applying a standard structural specification. In practice, for the regressions reported here, choices of instruments that passed overidentification tests generated results that were very similar to those reported. In some cases, the alternative set of instruments failed other specification tests, indicating that the instruments were weak.

The estimated differences in *mfp* by data source are reduced by the use of more flexible functional forms, methods that account for firm heterogeneity, and gross output production functions. The differences are not, however, completely accounted for. The potential explanations listed in section 6.1 (incorrect functional form, firm heterogeneity, or data inconsistencies) could still affect estimated *mfp*. Whether including a separate AES dummy in production regressions is an appropriate response to these differences, and how the coefficient on that dummy is interpreted, depends on beliefs about the source of the remaining differences.

Table 10: Translog production functions (industry EE11): value added and gross output

	Value Added			Gross Output		
	OLS	FE	ACF	OLS	FE	ACF
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Labour)	1.386***	1.341***	1.989***	1.433***	1.283***	1.746***
	[0.077]	[0.064]	[0.414]	[0.055]	[0.047]	[0.172]
ln(Capital)	-0.357***	-0.135**	-0.728**	0.114***	0.156***	0.467***
	[0.053]	[0.058]	[0.293]	[0.039]	[0.036]	[0.152]
ln(Intermed)				-0.691***	-0.623***	-1.236***
				[0.031]	[0.037]	[0.176]
ln(L)*ln(L)	0.068***	0.068***	0.104***	0.044***	0.039***	0.064***
	[0.007]	[0.006]	[0.030]	[0.004]	[0.003]	[0.012]
ln(K)*ln(K)	0.036***	0.022***	0.055***	0.008***	0.010***	0.005
	[0.003]	[0.003]	[0.016]	[0.002]	[0.002]	[0.004]
ln(M)*ln(M)				0.065***	0.064***	0.097***
				[0.002]	[0.002]	[0.010]
ln(K)*ln(L)	-0.083***	-0.081***	-0.141***	0.002	-0.001	0.028*
	[0.008]	[0.007]	[0.042]	[0.005]	[0.004]	[0.016]
ln(K)*ln(M)				-0.018***	-0.023***	-0.043***
				[0.003]	[0.003]	[0.010]
ln(L)*ln(M)				-0.101***	-0.088***	-0.147***
				[0.004]	[0.003]	[0.017]
AES	**0.347*****	**0.203*****	**0.161*****	**0.057*****	**0.058*****	**0.052*****
	[0.020]	**[0.017]**	**[0.011]**	**[0.007]**	**[0.006]**	**[0.003]**
Observations	49,779	49,779	33,534	50,970	50,970	34,164
Number of firms		12,357			12,567	
R-squared	0.70	0.41	0.73	0.95	0.87	0.96
RTS	0.965***	0.882***	1.017	0.986***	0.980***	1.003
	[0.007]	[0.010]	[0.017]	[0.004]	[0.005]	[0.007]
K-OutputElast	0.286***	0.229***	0.208***	0.046***	0.059***	0.050***
	[0.006]	[0.008]	[0.012]	[0.003]	[0.003]	[0.019]
L-OutputElast	0.679***	0.653***	0.808***	0.256***	0.237***	0.281***
	[0.008]	0.010	[0.023]	[0.006]	[0.006]	[0.016]
M-OutputElast				0.684***	0.684***	0.673***
				[0.003]	[0.004]	[0.013]
Allen-KL	-1.776	2.054**	3.527***	0.000***	0.000***	0.000***
	[1.401]	[1.017]	[1.051]	[0.000]	[0.000]	[0.000]
Allen-LM				-0.489***	-0.451***	-0.876***
				[0.000]	[0.000]	[0.012]
Allen-KM				0.000***	0.000***	-0.001***
				[0.000]	[0.000]	[0.000]
Hansen J			13.79			44.23
df(J)			22			18
p-value			0.909			0.000535
Bootstrap reps			50			49

*Note: includes constant, year effects, *** p<0.01, ** p<0.05, * p<0.1. (1): RTS=Returns to scale. Significance indicators reflect difference from 1 (ie, constant returns to scale).*

6.2. Does production function specification matter?

The discussion in the previous section, which was focused primarily on examining differences in *mfp* estimates on the basis of data source, has also highlighted the impact of different specifications and estimation methods. In this section we summarise the impacts on *mfp* dispersion, and on key production function parameters and elasticities.

6.2.1. Does the specification matter for *mfp* estimation?

As was evident in Figure 2, the dispersion of estimated *mfp* is larger when estimated from a value added production function than it is when estimated from a gross output production function. Table 11 provides estimates of the standard deviation and variance of *mfp* from selected specifications, for a consistent sub-sample of observations for which all measures are available.[44] The variance of value added *mfp* is roughly 4 to 5 times as large as the corresponding variance based on a gross output production function. The *mfp* distributions (based on all available observations) are shown graphically in Figure 3. The gross output *mfp* densities are considerably more peaked than those from value added production functions. Furthermore, dispersion of *mfp* from translog specifications is smaller than that from Cobb-Douglas specifications, particularly for gross output production functions, where the translog specification captures interactions between intermediates and other inputs which would otherwise be picked up in the residual.

Table 11: Productivity dispersion (EE11)

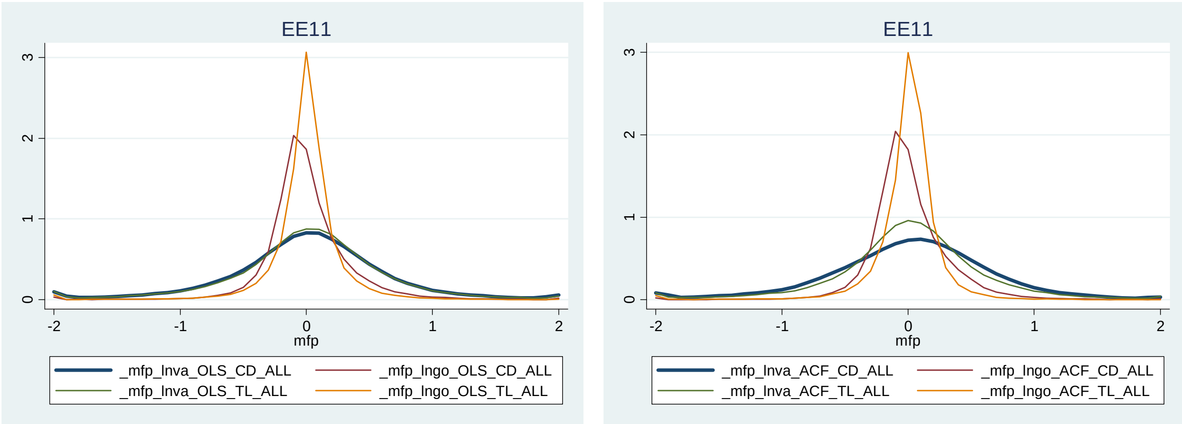
			Observations	sd	Variance
Value Added	Cobb-Douglas	OLS	33,534	0.671	0.450
		FE	33,534	0.707	0.500
		ACF	33,534	0.721	0.520
	Translog	OLS	33,534	0.650	0.422
		FE	33,534	0.668	0.447
		ACF	33,534	0.657	0.432
Gross Output	Cobb-Douglas	OLS	33,534	0.295	0.087
		FE	33,534	0.302	0.091
		ACF	33,534	0.301	0.090
	Translog	OLS	33,534	0.223	0.050
		FE	33,534	0.225	0.051
		ACF	33,534	0.230	0.053

Note: Statistics are for a sub-sample of observations for which all measures are available. The most restrictive specification is the value added ACF specification.

[44] The value added ACF estimates constrain the available observations because of the use of lags and the exclusion of firms with zero or negative value added. The sample restriction affects the relative size of OLS and ACF specifications but does not alter the broad patterns of difference by production function (VA/GO and CD/TL).

Figure 3: (kernel) Density plots for selected *mfp* estimates (EE11)

OLS *ACF*



Different *mfp* measures are highly correlated across different production function specifications (Cobb-Douglas and Translog) and across different estimation methods (OLS, FE, and ACF). Table 12 shows correlations separately for value added production function and gross output production functions. For specifications using the same production function, correlations range from 0.919 to 0.994. Correlations are lower when comparing across Cobb-Douglas and translog production functions (0.578 to 0.968). The correlation of Cobb-Douglas and translog specifications is lowest when using a gross output production function, suggesting that there may be significant factor interactions when intermediate consumption is included as a factor.

If we compare *mfp* estimates from value added production functions with those from gross output production functions (not shown), the correlation drops to between 60 and 70 percent.

Table 12: Correlation matrix for *mfp* estimates (EE11)

Value Added production function

		Cobb-Douglas			Translog		
		OLS	Firm FE	ACF	OLS	Firm FE	ACF
Cobb-Douglas	OLS	1.000					
	Firm FE	0.959	1.000				
	ACF	0.949	0.966	1.000			
Translog	OLS	0.962	0.917	0.906	1.000		
	Firm FE	0.953	0.968	0.932	0.974	1.000	
	ACF	0.926	0.902	0.869	0.988	0.978	1.000

Gross Output production function

		Cobb-Douglas			Translog		
		OLS	Firm FE	ACF	OLS	Firm FE	ACF
Cobb-Douglas	OLS	1.000					
	Firm FE	0.977	1.000				
	ACF	0.969	0.963	1.000			
Translog	OLS	0.809	0.784	0.785	1.000		
	Firm FE	0.835	0.819	0.805	0.994	1.000	
	ACF	0.582	0.578	0.578	0.933	0.919	1.000

Note: All correlations are calculated using a consistent sample, for which all measures are available.

Studies that examine productivity dynamics and change rely on robust estimation of the change in *mfp*. For such studies, high cross-sectional correlations do not ensure robust results. Table 13 shows correlations among annual changes in different estimates of *mfp*. The correlations are high for different estimates of value added production functions (0.939 to 0.994), regardless of functional form or estimation method. For gross output production functions, there is a lower correlation of annual changes (0.661 to 0.983).[45]

Estimates of annual *mfp* changes from value added production functions are less highly correlated with those based on gross output production functions (correlations of 0.617 to 0.734), reflecting differences in the impact of variable intermediate inputs.

Table 13: Correlation matrix for estimates of firm-level mfp growth (EE11)

			Value Added				Gross Output			
			Cobb-Douglas		Translog		Cobb-Douglas		Translog	
			OLS	ACF	OLS	ACF	OLS	ACF	OLS	ACF
Value Added	Cobb-Douglas	OLS	1							
		ACF	0.978	1						
	Translog	OLS	0.977	0.967	1					
		ACF	0.952	0.939	0.994	1				
Gross Output	Cobb-Douglas	OLS	0.686	0.665	0.644	0.617	1			
		ACF	0.714	0.665	0.662	0.636	0.983	1		
	Translog	OLS	0.721	0.703	0.734	0.732	0.841	0.832	1	
		ACF	0.678	0.673	0.706	0.708	0.662	0.661	0.953	1

6.2.2. Does the specification matter for parameter estimates?

While a good deal of economic research on productivity focuses on *mfp* patterns and changes, production function estimation provides insights into production technologies as well. The choice of production function as well as the estimation method can affect the nature and robustness of insights. As noted in section 2.2.1, simple specifications may constrain what is estimated (for instance, Cobb-Douglas specifications constrain output elasticities to be constant, and elasticities of substitution to equal one; value added specifications do not allow for intermediate inputs to interact with other inputs in production). Different estimation methods have developed to counteract various forms of bias in estimated production parameters (section 2.2.2).

In this section, we provide some general observations on the different estimates shown in the regression results in Table 9 and Table 10. These regressions show a range of estimates for output elasticities $\left(\frac{\partial lny}{\partial lnx}\right)$. For Cobb-Douglas production functions in Table 9, the output elasticity is the coefficient on the corresponding factor. For the translog production functions in Table 10, the output elasticity of a factor varies with the quantity of all factors. For these specifications, the estimated output elasticity is labelled as *OutputElas*, and is evaluated at the sample averages of all factors.

45 The low correlation of changes from the ACF specification may reflect specification errors related to invalid instruments used in estimation, as described in an earlier footnote.

The estimates of output elasticities are most sensitive to specification choices when output is measured as value added, and when production is modelled as a Cobb-Douglas production function. We focus on the 'pooled' estimates in columns 3, 6 and 9 of Table 9. The value added estimates of output elasticities of capital and labour are 0.316 and 0.603 respectively (column 3). The fixed effects estimates of both elasticities are lower (0.249 and 0.534), possibly reflecting attenuation bias due to measurement error (Griliches & Mairesse, 1998). The downward bias in the fixed effect estimates is also evident in the low estimated returns to scale (0.784), compared with the value of 0.919 estimated by *OLS*. Controlling for simultaneity bias using the *ACF* approach (column 9) reduces the coefficient on the more variable factor, labour. The output elasticity of capital is raised to 0.405 – almost as large as the labour elasticity (0.412). The returns to scale estimate from the *ACF* specification (0.816) is low relative to the *OLS* estimate, though the *ACF* estimates have large standard errors on the coefficients, and fail specification tests.

Estimates of output elasticities are more stable across specifications when using a gross output production specification. The reduction in coefficients is modest when fixed effects are used, and the *ACF* estimates show only a relatively small decline in the coefficient on the most variable factor – in this case, intermediates. Estimated returns to scale from gross output Cobb-Douglas production functions are higher than in the value added case. Estimates are 0.949 and 0.980 in *OLS* and *ACF* specifications, and are somewhat lower (0.917) in the fixed effects case, again suggesting attenuation bias.

Estimated standard errors on coefficients are again larger for the *ACF* specification than for *OLS* or *FE*, though the difference is smaller than in the value added case. In contrast with the value added case, the gross output *ACF* specification passes overidentification tests.

Estimated value added output elasticities continue to be more sensitive to specification than gross output elasticities, even when estimated with a more flexible functional form (translog). The first three columns of Table 10 show value added estimates of output elasticities ranging from 0.208 (*ACF*) to 0.286 (*OLS*) for capital, and from 0.653 (*FE*) to 0.808 (*ACF*) for labour. As in the Cobb-Douglas estimates, fixed effects estimates are smaller than *OLS* estimates. Estimated returns to scale drop from 0.965 for OLS to 0.882 for *FE*. The *ACF* estimate of 1.017 is not significantly different from 1 (constant returns to scale).

The more stable gross output elasticity estimates are shown in the final three columns of Table 10, ranging from 0.046 to 0.059 for capital, from 0.237 to 0.281 for labour, and from 0.673 to 0.684 for intermediates. Estimated returns to scale from the *FE* specification, while still lower than the *OLS* estimates, are not significantly so (0.980 compared with 0.986). Again, the *ACF* estimate is consistent with constant returns to scale (1.003).

Finally, one advantage of the translog specification is that it allows for variable output elasticities and elasticities of substitution. From standard production theory, we would expect diminishing returns to each factor, implying a negative coefficient on squared terms (ln(L)*ln(L), ln(K)*ln(K) and ln(M)*ln(M)). In Table 10, all of these estimates are positive for industry EE11, which is inconsistent with simple production theory. Estimates of the Allen elasticity of substitution are presented in the table. This elasticity measures the proportional change in the quantity of an input demanded when the price of another input changes proportionally (Allen, 1934; Stern, 2011). In the value added case, the elasticity of substitution between capital and labour is large and positive, implying that the inputs are substitutes (an increase in the price of one factor leads to an increase in the quantity of the other). In contrast, allowing for intermediates to be substitutable for other inputs produces estimates that imply virtually no interaction between capital and other inputs, but a complementarity between intermediates and labour. The estimated elasticity is negative, implying that an increased price of intermediates leads to a reduction in the use of labour.

6.3. Augmented production functions and geography

It is common in studies of productivity to augment productivity regressions by including additional regression covariates to estimate the contribution of other measured characteristics to *mfp* or *mfp* growth. Added variables may include firm-level characteristics such as foreign ownership, exporting status, the presence of immigrants, or R&D expenditure, or they may include industry or geographic variables to gauge the impact of competition or local employment density. The *LBD* provides many opportunities for such studies. The database is linked to business surveys such as the Business Operations Survey, which contains a breadth of information for a stratified sample of between 5,000 and 6,000 enterprises each year.[46]

To illustrate the estimation of augmented production functions, we present estimates of the productivity premium associated with operating in Auckland. This example also highlights different approaches to using geographic measures within the *LBD*.

6.3.1. Geographic analysis using the *LBD*

*PENT*s can operate in more than one location (as captured by its component plants, represented by *permanent business numbers*, or *PBN*s) but production information is available only at the *PENT* level. The first step in spatial analysis of productivity within the *LBD* is to link spatial information to the productivity dataset. Subsequently, there are alternative ways of modelling spatial productivity, to accommodate the aggregation of production data to the *PENT* level.

Data construction

It is necessary to extract information about the location of each *PENT*'s component *PBN*s. We also need information on the relative size of each *PBN* within the *PENT*. Production information is not available at the *PBN* level, so we rely on employment shares as the only indicator of relative size.

Within the *LBD*, the relationship between *PBN*s and enterprises is contained within the table *load_lbf_fact_business*. Plant locations are identified from the version of the *LBF* that is captured in the *IDI*. To maintain consistency with the production data measures of labour input, we link monthly PBN employment to PENTs using the *IDI* table *IDI_Sandpit.clean_read_IR.pent_emp_mth_FTE_IDI_20141205_RFabling*. Broad location information (territorial authority and regional council) can be linked in the *IDI* from *br_clean.pbn*, with more detailed location information available in the *LBD* table *load_lbf_fact_business*. Aggregating to consistent annual frequency is done using the following two correspondence tables stored in the *IDI* Sandpit: *clean_read_IR.pent_bal_date_IDI_20141205* and *clean_read_IR.dim_bal_date_year_IDI_20141205*.

In order to streamline the derivation, we have generated a table of monthly labour inputs for each PBN, with *PENT* identifiers. The table is stored in *IBULDD_RESEARCH_DATALAB* as

[46] Further information on the *BOS* is available in Fabling (2009) and from the Statistics New Zealand website http://www.stats.govt.nz/browse_for_stats/businesses/business_growth_and_innovation.aspx.

pent_pbn_month_L_IDI_20141205. This can be readily linked to *IBULDD* location information (*load_lbf_fact_business*) to obtain measures of employment shares by location for each *PENT*.

The production and location data can be linked in various ways, depending on the chosen approach to estimation.[47] The estimates presented below are based on a dataset with annual *PENT* observations. Location patterns are summarised by employment shares. Specifically, we use a variable that measures the proportion of a *PENT's* annual *FTE* employment accounted for by employment in the Auckland region. For single-location *PENT*s, this will take the values of zero or one.

Econometric analysis

We wish to estimate the relationship between a *PENT*'s productivity and the proportion of its *FTE* employment that is in Auckland (denoted $\tilde{A}$). In its most general form, an augmented production function can include the additional variable as an additional factor of production, potentially allowing it to interact with other factors in production.[48]

$$y = f\left(\tilde{A}, k, l, m; \beta\right) + e \tag{15}$$

A simpler, and commonly used specification is to augment the standard logged production function with an additive linear term, as in equation 16. The coefficient γ provides an estimate of the productivity premium associated with operating in Auckland:

$$y = \alpha + \gamma\tilde{A} + f(k, l, m; \beta) + e \tag{16}$$

In order to obtain a single estimate of γ across all firms, we would ideally include industry-specific production function coefficients (β), or estimate a system of industry-specific regressions, constraining γ to be equal in each equation. A commonly used expedient approach is to carry out estimation in two stages. In the first stage, an un-augmented production function is estimated, and an estimate of *mfp* ($\hat{e}$) is obtained. In the second stage, the estimated *mfp* is regressed on the augmenting variable. This approach is shown in equation 17.[49]

$$\begin{gathered} y = \alpha_1 + f(k, l, m; \beta) + e \\ \hat{e} = \alpha_2 + \gamma_2\tilde{A} + u \end{gathered} \tag{17}$$

The one stage (γ) and two-stage (γ_2) approaches will in general give different estimates but in practice they are often very close. Differences can arise due to correlation between $\tilde{A}$ and factor inputs due to, among other things, differences in factor prices across regions.

---

47 Options include allocating *PENT* production data to *PBN*s in proportion to employment, replicating *PENT*-level measures for each associated *PBN* observation, and weighting *PBN* observations by their share of the corresponding *PENT*. See Maré (2008) for an example and discussion of potential biases from different approaches. Estimating productivity using *PENT*-year observations and including spatial variables as employment-weighted means across component *PBN*s is our preferred approach.

48 See Graham & Kim (2007) or Maré & Graham (2009) for examples including a spatial variable (employment density) as an interacting factor of production.

49 Where the augmenting variable is a categorical variable, or does not vary within categories, the second stage can be estimated using category-level averages (eg: regressing mean *mfp* by city on employment density in each city).

In this section, we examine the sensitivity of estimates of γ to different production function specifications, and different estimation approaches – including two-stage estimation. Given the relatively high correlation of mfp estimates across specifications, it is possible that estimates of γ may be relatively similar across specifications. Whether this is the case will, of course, depend on the augmenting variable, and the sample of firms. The findings that we present should be interpreted as case specific – applying only to the effect of a firm's presence in Auckland for productivity within the construction industry (EE11).

Within the construction industry between 2001-2012, 22% of *PENT*s had employees only in Auckland. A further 3% employed both in the Auckland region and in another region. The *PENT*s that operated in Auckland and elsewhere were on average larger, accounting for 18% of industry employment, compared with 16% accounted for by Auckland-only *PENT*s.

Table 14 summarises estimates of the Auckland productivity premium for industry EE11. Each coefficient is from a different regression. The first row reports the mean difference in value added per worker (VAPW) between Auckland based and other *PENT*s (from a regression of the log of value added per worker on $\tilde{A}$). The estimate of 0.253 implies that Auckland firms have $VAPW$ that is around 29% higher ($e^{0.253} - 1$) than that of firms outside Auckland. The estimate in the second row is from a regression of ln(value added) on ln(FTE) – a value added production function with labour as the only measured input. The estimate of 0.193 is lower, and the coefficient on ln(FTE) (not shown) of 0.78 is significantly lower than 1.[50]

Controlling for capital inputs using a value added production function further lowers the estimated Auckland premium, to 0.14, whether estimated with a Cobb-Douglas or with a translog production function. Auckland firms are, on average, more capital intensive than non-Auckland firms. When we incorporate intermediate inputs in a gross output production function, the estimated Auckland premium drops further – to 0.066 when we use a Cobb-Douglas production function. It drops further, to 0.037, when we allow for more flexible productive interactions between factors with a translog production function.

[50] This is higher than the estimate of the Auckland *VAPW* premium in construction of 0.144 implied by the estimate in Appendix 7 of Maré (2008), which used a different industry definition (ANZSIC96 industry E41) over a different time period (2000-2006). This earlier estimate used a combination of *AES* and *BAI* data, rather than the more robust productivity dataset documented in the current paper.

Table 14: The Auckland productivity premium: Augmented production function (EE11)

		OLS	OLS (two stage)	FE	ACF
VAPW		0.253***			
		[0.015]			
VA		0.193***			
		[0.015]			
Value Added	Cobb-Douglas	0.141***	0.138***	0.070	0.069***
		[0.013]	[0.013]	[0.073]	[0.011]
	Translog	0.142***	0.139***	0.057	0.069***
		[0.013]	[0.012]	[0.074]	[0.011]
Gross Output	Cobb-Douglas	0.066***	0.064***	0.034	0.038***
		[0.006]	[0.006]	[0.030]	[0.004]
	Translog	0.037***	0.036***	0.027	0.038***
		[0.005]	[0.005]	[0.030]	[0.004]

Note: Each entry from a separate regression. Robust standard errors are clustered by PENT.

The second column of Table 14 presents the two-stage estimates of γ_2. These are very close to the single-stage estimates in the first column. Although there is no advantage to using a two-stage approach when looking at a single industry, the similarity of results in the first two columns is encouraging for the use of the two-stage approach.[51]

Fixed effects estimators based on value added production functions are much smaller than the corresponding *OLS* estimates, consistent with Auckland firms having permanently higher productivity for reasons possibly unrelated to being in Auckland. The standard errors are also considerable larger, which is unsurprising given that, in the presence of firm fixed effects, the coefficient on $\tilde{A}$ is identified from within-firm changes in $\tilde{A}$ over time. For 97% of *PENT*s (accounting for 82% of employment), there is no variation over time in $\tilde{A}$ – they always have their employment entirely within Auckland, or entirely outside Auckland. Controlling for the use of intermediates, through a gross output production function, produces smaller estimates of the Auckland productivity premium, though still with large standard errors, so that the estimates are not statistically different from zero. Two-stage estimates of the fixed effects specification, with firm fixed effects included in both the first and second stages (and with the estimated first stage fixed effects included in *mfp*) are almost identical to the one-stage estimates, and are not reported here.

Controlling for firm heterogeneity using the approach of Ackerberg, Caves & Frazer yields similar estimates to those obtained from fixed effects estimation, though the estimates are more precisely estimated because they gain some identification from immobile firms. Value added production function estimates imply a premium of 6.9%, whereas the estimates from the gross output specification are 3.8%.

In this application, the use of one-stage or two-stage estimation does not produce markedly different estimates, but the choice of value added as opposed to gross output production

[51] The standard errors reported in the two-stage specification have not been adjusted for the fact that the dependent variable has been estimated. Correct standard errors could be obtained by bootstrapping the first and second stages.

functions does matter, with a smaller premium estimated when intermediates are allowed to enter the production function. In *OLS* estimates, using a more flexible functional form captures more of the Auckland effect within the production function, leaving a smaller *mfp* premium. The *FE* and *ACF* specifications absorb a similar proportion of the premium.

What is being identified as a (possibly firm-specific) *mfp* premium in 'simpler' specifications appears to be related to the interacting effects of labour and materials in production. In gross output specifications that accommodate this, the estimated *mfp* premium is reasonably similar across specifications, ranging from 0.027 to 0.038. This is a much more modest premium than the 25% raw difference in labour productivity (*VAPW*).

7. Concluding comments

This paper has documented the methods used to create a productivity dataset that is available for use by researchers within Statistics New Zealand's secure microdata environment. The dataset combines information from survey and administrative data sources in a consistent way. Inevitably, some modelling, imputation and adjustments have been applied. The data are almost certainly imperfect and incomplete. They are not a substitute for aggregate production and productivity statistics but do provide a sound basis for estimating the performance of firms. The ability to link the productivity dataset with other data in the *LBD* and *IDI* provides many avenues for future research.

We have highlighted some of the key issues in using the dataset for productivity estimation. Not only are there important (functional form) choices to be made about how to model the production process there are also long-standing challenges of identifying and interpreting underlying production parameters. We have provided a selection of estimates using data on one industry – *EE11: Building Construction* – to illustrate the sensitivity of estimates to alternative modelling and estimation choices.

The empirical analysis is illustrative rather than comprehensive, but has clearly demonstrated the ability of the *LBD* and the productivity dataset to support meaningful analysis. The preparation of the productivity dataset, and this paper's discussion of data and methods will hopefully prove to be a useful resource for researchers who are carrying out research on New Zealand firms.

8. References

Abramovitz, M. (1956). Resource and Output Trends in the United States Since 1870. *The American Economic Review*, 5–23.

Ackerberg, D. A., Caves, K., & Frazer, G. (2006). Structural identification of production functions. *Mimeo, UCLA, December 28, 2006*. Retrieved from _http://www.econ.ucla.edu/ackerber/ACF20withtables.pdf_

Allen, R. G. (1934). A comparison between different definitions of complementary and competitive goods. *Econometrica: Journal of the Econometric Society*, 168–175.

Berndt, E. R., & Khaled, M. S. (1979). Parametric productivity measurement and choice among flexible functional forms. *The Journal of Political Economy*, 1220–1245.

Cobbold, T. (2003). *A comparison of gross output and value added methods of productivity estimation* (Research memorandum). Canberra: Australian Government Productivity Commission.

Conway, P., & Zheng, G. (2014). *Trade over distance for New Zealand firms: Measurement and implications* (New Zealand Productivity Commission Working Paper No. 2014/5). Wellington.

Devine, H., Doan, T., Iyer, K., Mok, P., & Stevens, P. (2012). *Decomposition of New Zealand firm productivity, 2001-2008* (Paper presented to NZAE conference). Wellington New Zealand.

Devine, H., Doan, T., & Stevens, P. (2012). *Explaining Productivity Distribution in New Zealand Industries: The effects of input quality on firm* (Paper presented to NZAE conference).

Doan, T., Maré, D., & Iyer, K. (2014). Productivity spillovers from foreign direct investment in New Zealand. *New Zealand Economic Papers*, 1–27. http://doi.org/10.1080/00779954.2014.945229

Fabling, R. (2009). *A rough guide to New Zealand's Longitudinal Business Database*. Tokyo: Hitotsubashi University.

Fabling, R. (2011). Keeping it together: Tracking firms in New Zealand's Longitudinal Business Database. *Motu Working Paper*, *11-01*.

Fabling, R., Gemmell, N., Kneller, R., & Sanderson, L. (2013). *Estimating firm-level effective marginal tax rates and the user cost of capital in New Zealand* (Motu Working Paper No. 13-14). Wellington: Motu Economic and Public Policy Research.

Fabling, R., & Grimes, A. (2014). The 'suite' smell of success: Personnel practices and firm performance. *ILR Review*, *67*(4), 1095–1126.

Fabling, R., Grimes, A., Sanderson, L., & Stevens, P. (2008). Some rise by sin, and some by virtue fall firm dynamics market structure and performance. *NZ Ministry of Economic Development Occasional Paper*, *08/01*.

Fabling, R., & Maré, D. C. (2015). Addressing the absence of hours information in linked employer-employee data. *Motu Working Paper*, *(forthcoming)*.

Fabling, R., & Sanderson, L. (2013). Exporting and firm performance: Market entry, investment and expansion. *Journal of International Economics*, *89*(2), 422–431.

Fabling, R., & Sanderson, L. (2014). *Productivity distributions in New Zealand: The dangers of international comparisons* ((mimeo)). Motu Research.

Fuss, M., McFadden, D. L., & Mundlak, Y. (1978). A survey of functional forms in the economic analysis of production. In M. Fuss & D. L. McFadden (Eds.), *Production economics: A dual approach to theory and applications*. Amsterdam: North Holland Publishing.

Giannakas, K., Tran, K. C., & Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, *28*(1), 75–100.

Graham, D. J., & Kim, H. Y. (2007). An empirical analytical framework for agglomeration. *Annals of Regional Science*, *2008*(42), 267–289.

Grieco, P. L., Li, S., & Zhang, H. (2013). *Production Function Estimation with Unobserved Input Price Dispersion*. Pennsylvania State University Working Paper.

Griffin, R. C., Montgomery, J. M., & Rister, M. E. (1987). Selecting functional form in production function analysis. *Western Journal of Agricultural Economics*, 216–227.

Griliches, Z., & Mairesse, J. (1998). Production functions: The search for identification. In S. Strøm (Ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium* (pp. 169–203). Cambridge: Cambridge University Press.

Grimes, A., Ren, C., & Stevens, P. (2009). The Need for Speed: Impacts of Internet Connectivity on Firm Productivity. *Journal of Productivity Analysis*, *37*(2), 187–201.

Jorgenson, D. W. (1986). Econometric methods for modeling producer behavior. In Griliches, Zvi & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 3, pp. 1841–1915). Amsterdam: North Holland.

Klette, T. J. (1999). Market power, scale economies and productivity: Estimates from a panel of establishment data. *Journal of Industrial Economics*, *47*(4), 451–476.

Klette, T. J., & Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Economics*, *11*(4), 343–361.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, *70*(2), 317–341.

Mai, B., & Warmke, N. (2012). Comparing approaches to compiling macro and micro productivity measures using Statistics New Zealand data. Presented at the New Zealand Association of Economists Conference, Palmerston North, New Zealand.

Maré, D. C. (2008). Labour productivity in Auckland firms. *Motu Working Paper*, *08-12*. Retrieved from _www.motu.org.nz_

Maré, D. C., & Fabling, R. (2013). Productivity and local workforce composition. In R. Crescenzi & M. Percoco (Eds.), *Geography, institutions and regional economic performance* (pp. 59–76). Berlin & Heidelberg: Springer-Verlak.

Maré, D. C., & Graham, D. J. (2009). Agglomeration elasticities in New Zealand. *Motu Working Paper*, *09-06*.

Maré, D. C., & Graham, D. J. (2013). Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, *75*(0), 44–56.

Maré, D. C., & Hyslop, D. R. (2006). *Worker-Firm Heterogeneity and Matching: An analysis using worker and firm fixed effects estimated from LEED*. Wellington: Statistics New Zealand.

Maré, D. C., & Timmins, J. (2006). Geographic concentration and firm productivity. *Motu Working Paper*, *2006-08*.

Martin, R. (2008). Productivity dispersion, competition and productivity measurement. *CEP Discussion Paper*, *692*.

Martin, R. (2010). *Productivity spreads, market power spreads and trade* (CEP Discussion Paper No. 997). Centre for Economic Performance.

Nolan, P. (2014). Lifting New Zealand's productivity: A research agenda. *Policy Quarterly*, *10*(2), 22–29.

OECD. (2001). *Measuring Productivity: OECD Manual - Measurement of aggregate and industry-level productivity growth*. Paris: OECD.

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*(6), 1263–1297.

Seyb, A. (2003). *The Longitudinal Business Frame*. Wellington: Statistics New Zealand. Retrieved from _www.stats.govt.nz_

Statistics New Zealand. (2014a). *Productivity Statistics (1978-2013) - Hot Off The Press*. Wellington: Statistics New Zealand.
Statistics New Zealand. (2014b). *Productivity Statistics: Sources and methods (10th Edition)*. Wellington: Statistics New Zealand.
Stern, D. I. (2011). Elasticities of substitution and complementarity. *Journal of Productivity Analysis*, *36*(1), 79–89. http://doi.org/10.1007/s11123-010-0203-1
Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, *49*(2), 326–365. http://doi.org/10.1257/jel.49.2.326
Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, *104*(3), 112–114. http://doi.org/10.1016/j.econlet.2009.04.026

9. Appendix 1: Industry groupings

Appendix Table 1: Industry Groupings (ANZSIC06/ NZSIOC)

PF_IND	PPI (NZSIOC level 3)	Description	Number of National Accounts Working Inds	ANZSIC06 industry codes	Measured sector
	AA	Agriculture, forestry, and fishing			
AA11	AA11	Horticulture and fruit growing	1	A011/ A012/ A013	Yes
AA12	AA12	Sheep, beef cattle, and grain farming	1	A014/ A015	Yes
AA13	AA13	Dairy cattle farming	1	A016	Yes
AA14	AA14	Poultry, deer, and other pstock farming	1	A017/ A018/ A019	Yes
AA21	AA21	Forestry and logging	1	A030	Yes
AA31	AA31	Fishing and aquaculture	1	A020/ A041	Yes
AA32	AA32	Agric, forest, fish support services, and hunting	1	A042/ A051/ A052	Yes
BB11	BB	Mining	1	B	Yes
	CC	Manufacturing			
CC1	CC11	Meat and meat product manufacturing	1	C111	Yes
CC1	CC12	Seafood processing	1	C112	Yes
CC1	CC13	Dairy product manufacturing	1	C113	Yes
CC1	CC14	Fruit, oil, cereal, and other food manufacturing	1	C114/ C115/ C116/ C117/ C118/ C119	Yes
CC1	CC15	Beverage and tobacco product manufacturing	1	C12	Yes
CC21	CC21	Textile, leather, cloth, and footwear manufacturing	1	C13	Yes
CC3	CC31	Wood product manufacturing	1	C14	Yes
CC3	CC32	Pulp, paper, and converted paper manufacturing	1	C15	Yes
CC41	CC41	Printing	1	C16	Yes
CC5	CC51	Petroleum and coal product manufacturing	1	C17	Yes
CC5	CC52	Basic chemical and chemical product manufacturing	3	C18	Yes
CC5	CC53	Polymer product and rubber product manufacturing	1	C19	Yes
CC61	CC61	Non-metallic mineral product manufacturing	1	C20	Yes
CC7	CC71	Primary metal and metal product manufacturing	1	C21	Yes
CC7	CC72	Fabricated metal product manufacturing	1	C22	Yes
CC81	CC81	Transport equipment manufacturing	1	C23	Yes
CC82	CC82	Machinery and other equipment manufacturing	2	C24	Yes
CC91	CC91	Furniture and other manufacturing	2	C25	Yes

PF_IND	PPI (NZSIOC level 3)	Description	Number of National Accounts Working Inds	ANZSIC06 industry codes	Measured sector
	DD	Electricity, gas, water, and waste services			
DD1	DD11	Electricity and gas supply	3	D26/ D27	Yes
DD1	DD12	Water, sewer, drainage, and waste services	3	D28/ D29	Yes
	EE	Construction			
EE11	EE11	Building construction	2	E30	Yes
EE12	EE12	Heavy and civil engineering construction	1	E31	Yes
EE13	EE13	Construction services	1	E32	Yes
FF11	FF	Wholesale trade	5	F	Yes
	GH	Retail trade and accommodation			
GH11	GH11	Motor vehicle & parts, and fuel retailing	2	G39/ G40	Yes
GH12	GH12	Supermarket, grocery, and specialised food retailing	2	G41	Yes
GH13	GH13	Other store-based and non-store retailing	4	G42/ G43	Yes
GH21	GH21	Accommodation and food services	2	H	Yes
	II	Transport, postal, and warehousing			
II11	II11	Road transport	1	I46	Yes
II12	II12	Rail, water, air, and other transport	3	I47/ I48/ I49/ I50	Yes
II13	II13	Post, courier support, and warehouse services	3	I51/ I52/ I53	Yes
	JJ	Information media and telecommunications			
JJ11	JJ11	Information media services	3	J54/ J55/ J56/ J57	Yes
JJ12	JJ12	Telecommunication, Internet, and library services	2	J58/ J59/ J60	Yes
	KK	Financial and insurance services			
KK1_	KK11	Finance	1	K62	Yes
KK1_	KK12	Insurance and superannuation funds	3	K63	Yes
KK13	KK13	Auxiliary finance and insurance services	1	K64	Yes
	LL	Rental, hiring, and real estate services			
LL11	LL11	Rental and hiring services	1	L66	Yes[a]
LL12	LL12	Property operators and real estate services	3	L67	Yes[ab]
	LL21	Ownership of owner-occupied dwellings	1		No
	MN	Professional and administrative services			
MN11	MN11	Professional, scientific, and tech services	5	M	Yes[a]
MN21	MN21	Administrative and support services	3	N	Yes[a]
		Local government administration	1	O753	No
		Central government admin, defence and public safety	3	O except for O753	No
		Education and training	4	P	No
		Health care and social assistance	3	Q	No
	RS	Arts, recreation, and other services			
RS11	RS11	Arts and recreation services	3	R	Yes
RS21	RS21	Other services	3	S	Yes[a]

Notes: (a) industry was formerly excluded from the measured sector. (b) We exclude the ANZSIC06 industry L67 from our analysis because the AES form for that industry does not collect balance sheet information.

Appendix Table 2: Industry Groupings (ANZSIC96)

pf_ind	ANZSIC96	Description	Measured Sector	Published	PPI
	A01	Agriculture	Yes	AA	
A011	A011	Horticulture and fruit growing	Yes		A01
A012	A012	Sheep and beef farming	Yes		A02
A013	A013	Dairy cattle farming	Yes		A03
A01_	A014-A016	Cropping and other farming	Yes		A04
A02	A02	Services to Agriculture; Hunting and Trapping	Yes	AA?	A05
A03	A03	Forestry and Logging	Yes	AB	A06
A04	A04	Commercial Fishing	Yes	AC	A07
B	B11	Coal Mining	Yes	BA	B01
B	B12	Oil and Gas Extraction	Yes	BA	B02
B	B13	Metal Ore Mining	Yes	BA	B03
B	B14	Other Mining	Yes	BA	B04
B	B15	Services to Mining	Yes	BA	B05
C21	C21	Food, Beverage and Tobacco	Yes	CA	
C21	C211	Meat and meat product Mfg	Yes		C01
C21	C212	Dairy product manufacturing	Yes		C02
C21	C213-C218	Other food manufacturing	Yes		C03
C21	C219	Tobacco, beverage and malt manufacturing	Yes		C04
C22	C22	Textile, Clothing, Footwear and Leather Manufacturing	Yes	CB	C06
C23	C23	Wood and Paper Product Manufacturing	Yes	CC	
C23	C231-C232	Wood product mfrg	Yes		C07
C23	C233	Paper and paper product mfrg	Yes		C08
C24	C24	Printing, Publishing and Recorded Media	Yes	CD	C09
C25	C25	Petroleum, Coal, Chemical and Associated Product Manufacturing	Yes	CE	
C25	C251-C253	Petroleum, coal and basic chemical mfrg	Yes		C10
C25	C254-C256	Rubber, plastic and other chemical mfrg	Yes		C11
C26	C26	Non-Metallic Mineral Product Manufacturing	Yes	CF	C12
C27	C27	Metal Product Manufacturing	Yes	CG	
C27	C271-C273	Basic metal mfrg	Yes		C13
C27	C274-C276	Sheet and fabricated metal mfrg	Yes		C14
	C28	Machinery and Equipment Manufacturing	Yes	CH	
C28a	C281-C282	Transport equipment mfrg	Yes		C15
C28b	C283-C286	Machinery and equipment mfrg	Yes		C16
C29	C29	Other Manufacturing	Yes	CI	C17
D	D36	Electricity and Gas Supply	Yes	DA	
D	D361	Electricity generation and supply	Yes		D01
D	D362	Gas supply	Yes		D02
D	D37	Water Supply, Sewerage and Drainage Services	Yes	DA	D03
E41	E41	General Construction	Yes	EA	E01
E42	E42	Construction Trade Services	Yes	EA	E01
F45	F45	Basic Material Wholesaling	Yes	FA	F01
F46	F46	Machinery and Motor Vehicle Wholesaling	Yes	FA	F01
F47	F47	Personal and Household Good Wholesaling	Yes	FA	F01
G51	G51	Food Retailing	Yes	GA	G01

pf_ind	ANZSIC96	Description	Measured Sector	Published	PPI
G52	G52	Personal and Household Good Retailing	Yes	GA	G01
G53	G53	Motor Vehicle Retailing and Services	Yes	GA	G01
H	H57	Accommodation, Cafes and Restaurants	Yes	HA	H01
I6a	I61	Road Transport	Yes	IA	I01
I6_	I62	Rail Transport	Yes	IA	I09
I6_	I63	Water Transport	Yes	IA	I03
I6_	I64	Air and Space Transport	Yes	IA	I04
I6_	I65	Other Transport	Yes	IA	I09
	I66	Services to Transport	Yes	IA	
I6a	I661	Services to Road Transport	Yes		I01
I6_	I662	Services to Water Transport	Yes		I03
I6_	I663	Services to Air Transport	Yes		I04
I6_	I664	Other Services to Transport	Yes		I09
I6_	I67	Storage	Yes	IA	I09
J	J71	Communication Services	Yes	JA	J01
K7_	K73	Finance	Yes	KA	K01
K7_	K74	Insurance	Yes	KA	K02
K75	K75	Services to Finance and Insurance	Yes	KA	K03
-	L77	Property Services	No		
-	L771,L772	Property operators, developers, real estate	No[1]	LA	L01
-	L773,L774	Non-financial asset investors, hiring and leasing	No	LB	L03
L78	L78	Business Services	Yes	LC	L04
-	M81	Government Administration	No		
-	All except M8113		No	MA	M01
-	M8113	Local government administration	No	MB	M02
-	M82	Defence	No	MA	M01
-	N84	Education	No	NA	N01
-	O86	Health Services	No	OA	O01
-	O87	Community Services	No	OA	O01
P	P91	Motion Picture, Radio and Television Services	Yes	PA	P01
P	P92	Libraries, Museums and the Arts	Yes	PA	P01
P	P93	Sport and Recreation	Yes	PA	P01
Q	Q95	Personal Services	Yes	QA	Q01
Q	Q96	Other Services	Yes	QA	Q01
Q	Q97	Private Households Employing Staff	Yes	QA	Q01
-	R99	Not Elsewhere Included	No	n/ a	n/ a

Property services included in the measured sector definition since 2011, with data back to 1996.

Appendix Table 3: Definition of commonly measured expenditure

AES Qn	AES variable		IR10 Qn	IR10 variable	
19	presale_amt	Purchases for resale			
18	purch_amt	Purchases	4	i10_prchases	purchases (including for resale)
35	purch_amt	all other operating exp (including rental and hiring; payment to subcontractors	14	i10_entertmt	entertainment
			18	i10_legalexp	legal
			19	i10_rates	Rates
			20	i10_rentcd	Rental and lease payments
			21	i10_repmaint	Repairs and maintenance
			22	i10_resdevcd	Research and development
			24	i10_subcntpts	Subcontractor payments
			25	i10_travacom	Travel and accommodation
			26	i10_vehexp	vehicles (excluding depreciation)
			27	i10_othexp	Other expenses
30	roypd_amt	Other royalties and patent fees			
56	amort_amt	Amortisation of intangible assets			
8	commtaxa_amt	Excise duties			
33+34	nctxfbt_amt	Road user charges, rates and other central and local government fees(includes Q33:commtaxb_amt (Road User Charges))			

Excluded items

31		Depreciation and amortisation	13		Depreciation
32		FBT	15		FBT
25		salaries and wages	23		salaries and wages
27		interest paid	17		Interest expenses
28		insurance	16		Insurance (exclude ACC)
36		Non-operating expenditure (incld bad debts . . .)	12		Bad debts written off
21		manufacturing and processing charges and fees			
23		Workplace injury insurance			
24		employer contributions			

10. Appendix 2(a): AES form

There are 28 different questionnaire forms for the AES. Different forms are administered to different industry groupings, to ensure that questions are relevant to the activities of the industry, and that the information collected provides relevant financial information.

The form shown here is for Manufacturing and Wholesale industries.

Sample forms for other industries are available from the Statistics New Zealand website:
http://www2.stats.govt.nz/domino/external/quest/sddquest.nsf/

Statistics New Zealand
TATAURANGA AOTEAROA

AF/MW/01

Annual Enterprise Survey 2012/2013
Manufacturing and Wholesale

For help and information:
Phone: 0800 333 105
64 9 920 9103
Email: aes@stats.govt.nz
Mail: Statistics New Zealand
Freepost 10007
Private Bag 92003
Auckland

Are the address details above correct? If not, please use the boxes below to correct any errors.

Attention

Legal name

Building / Level / Unit

Street or PO Box

Suburb / District or Rural Delivery

City / Town   Postcode

Please complete, sign and return this questionnaire in the envelope supplied, by

Purpose of this survey

The purpose of this survey is to collect statistics from a cross section of New Zealand businesses in a wide selection of industries.

The financial data collected by this survey is used in the production of a range of economic indicators. An example is Gross Domestic Product (GDP), which measures economic activity in New Zealand. These statistics help government and other organisations in planning and decision making.

For the purpose of this survey, a 'business' can be an enterprise or an organisation that operates for profit or not for profit (eg shop, company, trust, club).

Compulsory requirement

This survey has been approved by the Minister of Statistics, and the return of this questionnaire duly filled in and signed is a compulsory requirement under the Statistics Act 1975.

Confidentiality of information supplied

Only people authorised by the Statistics Act 1975 are allowed to see your individual information, and they must use it only for statistical purposes. Your information will be combined with similar information to prepare summary statistics.

As Acting Government Statistician I thank you for completing this survey. Your information contributes to statistics available for business decision-making. To find out how Statistics New Zealand can help your business grow, contact our information centre on 0508 525 525

Dallas Welch
Acting Government Statistician

Annual Enterprise Survey

Sample Only

Please read this first

1 Only include information for the business named on the front page. Do not provide consolidated data.

Don't include:
- subsidiary or associated businesses
- accounting divisions that operate entirely outside New Zealand

2 How to answer
- use a blue or a black pen
- mark your answers like this:
- if you make a mistake do this:
- print in capital letters like this: J O N E S. Write numbers like this: 2 6 0.
- where actual amounts are not available please give careful estimates.
- if any numbers **are negative**, show them with **a minus** sign like this: $ - 3 000

Round numbers up or down to the nearest thousand, as below:
- for example, if your answer is $127,138, round this to $127,000 and fill in the boxes like this: $ 1 2 7 000
- if your answer is $683, round this to $1,000 and fill in the boxes like this: $ 1 000
- if your answer is less than $500, or the question does not apply to this business, leave the boxes blank. $ 000

3 Accounts

To assist processing staff in interpreting your responses, please send with the completed questionnaire a copy of your accounts for the period covered (ie statement of financial performance, statement of financial position, depreciation schedule and the accompanying notes to the accounts).

4 Please keep a record of the time it takes you to complete this questionnaire. You are asked to record this at the end of the questionnaire.

Include:
- the time spent reading the instructions, working on the questions and obtaining information
- the time spent by all employees in collecting and providing this information

Information required

5 Please provide financial data for a balance date of

If your business uses a different balance date, please correct this here: Day Month Year

Note: the financial data you provide should have a balance date (ie last day of the financial year) between 1 October 2012 - 30 September 2013.

6 Does the financial data you will use cover a 12 month period?

- yes → Go to **7**
- no → The period covered is Day Month Year to Day Month Year

Please mark a reason why it is not a 12 month period.
- new business
- ceased during the year
- other → please specify

7 Please supply GST exclusive amounts if possible.

The amounts given in this questionnaire will:
- exclude GST
- include GST

Excise duties

8 Excise duties paid on the following goods manufactured in New Zealand: alcohol, tobacco products, motor vehicles, fuels

Don't include excise duties anywhere else in the questionnaire. $ 000 0030

Income

9 Sales of goods purchased for resale without further processing (ie sold in the same form in which they were purchased) $ 000 0120

Please describe the main items

10 Sales of goods manufactured or processed from stocks owned by this business $ 000 0150

Please describe the main items

Remember to round your answers to the nearest thousand dollars. See question **2** on page 2.

11 Income from services provided by this business

Include items such as:
- manufacturing and processing fees, charges (eg cut, make and trim charges, livestock slaughtering charges, wool scouring charges)
- repair services

Don't include:
- finance leasing
- renting and leasing of plant and machinery

$ 000 0200

Please describe the main items

12 For the amounts given in questions **9** to **11**, please describe any significant differences between this financial year and the previous financial year (eg increase in sales due to new product line)

13 Interest received

Include items such as:
- resident withholding tax on interest
- interest from banks, affiliated companies, and from finance leases
- use-of-money interest from the Inland Revenue Department (IRD)

$ 000 0501

14 Dividends received

Include:
- imputation credits
- resident withholding tax on dividends

$ 000 0503

15 Income from direct government subsidies and cash grants

Include items such as employment subsidies.

$ 000 0700

Please describe the main items

16 All other operating income

Include items such as:
- management fees
- royalties and patent fees
- income from operating leases
- renting and leasing of plant and machinery

Don't include items already recorded in the income section above.

$ 000 0801

Please describe the main items

17 Non-operating items of income

Include items such as:
- bad debts recovered
- profit on sale of capital assets above book value
- gains on revaluation of capital assets
- gains on exchange transactions
- insurance claims received
- gains before tax from extraordinary transactions
- subvention receipts

$ 000 0915

Please describe the main items

18 TOTAL INCOME
Total of questions 9 to 17

$ 000 1000

This is total A

Expenditure

19 Goods purchased for resale without further processing
(ie for sale in the same form in which they were purchased)

Include import duty paid on these purchases.
Don't include materials for use in processing and packaging.

$ 000 1120

Please describe the main items

20 Purchases of other goods and materials

Include items such as:
- materials for use in manufacturing
- packaging materials and containers

Don't include goods purchased for resale.

$ 000 1150

Please describe the main items

21 Manufacturing and processing charges and fees paid to other businesses or other accounting divisions of this business

Include items such as:
- cut, make and trim charges
- livestock slaughtering charges
- wool scouring charges

$ 000 1210

Please describe the main items

22 For the amounts given in questions 19 to 21 please describe any significant differences between this financial year and the previous financial year
(eg increase in expenditure due to a lower exchange rate)

23 Workplace injury insurance expenses

Include items such as:
- employer levies paid to Accident Compensation Corporation (ACC)
- workplace injury insurance premiums

On behalf of working proprietors and partners	+	Paid employees	=	Total payments to insurers/ACC
$ 000 (2601)		$ 000 (2611)		$ 000 (1401)

24 Employer contributions to superannuation, pension, and welfare schemes

Include items such as:
- withholding tax on contributions to superannuation funds
- payments for healthcare and medical insurance
- KiwiSaver contributions

On behalf of working proprietors and partners	+	Paid employees	=	Total employer contributions
$ 000 (2602)		$ 000 (2612)		$ 000 (1402)

25 Salaries and wages paid

Include items such as:
- severance and redundancy payments
- sick and holiday pay
- directors' fees
- bonus and other performance payments (for example commissions)
- other employment related expenses

Don't include:
- capitalised salaries and wages
- fringe benefit tax
- payments to contractors

Paid to working proprietors and partners	+	Paid employees	=	Total salaries and wages paid
$ 000 (2603)		$ 000 (2613)		$ 000 (1403)

26 Of the total salaries and wages paid in question **25**, please estimate the amount spent on severance and redundancy payments.
If this does not apply to this business, leave blank.

$ 000 (2649)

27 Interest paid

Include items such as:
- interest paid to banks, affiliated companies, and on finance leases
- use-of-money interest to the Inland Revenue Department (IRD)

$ 000 (1501)

28 Business insurance premiums $ 000 (1813)

29 Software licence fees that have not been capitalised $ 000 (1503)

30 Other royalties and patent fees paid

Include items such as:
- payments for the use of trademarks, brands and copyrights
- franchise lease payments
- Annual Catch Entitlement (ACE)

$ 000 (1504)

Please describe the main items

31 Depreciation and amortisation expense

Include items such as:
- depreciation on finance lease assets operated by this business
- amortisation of intangible assets

$ 000 (1602)

32 Fringe Benefit Tax (FBT) paid

Don't include withholding tax paid on contributions to superannuation funds. $ 000 (1700)

33 Road User Charges $ 000 (1703)

34 Rates and other central and local government fees

Don't include:
- withholding tax paid on contributions to superannuation funds
- GST, FBT and excise duties
- road user charges

$ 000 (1740)

35 All other operating expenses paid to other businesses

Include items such as:
- contract freight
- rental and hiring
- payments to contractors

Don't include items already recorded in the expenditure section above.

$ 000 (1841)

Please describe the main items

36 Non-operating items of expenditure

Include items such as:
- bad debts written off
- loss on sale of capital assets below book value
- loss on revaluation of capital assets
- loss on exchange transactions
- losses before tax from extraordinary transactions
- subvention payments

$ 000 (1915)

Please describe the main items

37 TOTAL EXPENDITURE
TOTAL of questions **19** to **36** excluding question **26**

$ 000 (2000)

This is total B

Stocks

38 Stocks of goods purchased for resale

Include items such as trading stocks.

Opening stocks		Closing stocks	
$ 000	2120	$ 000	2220

39 Stocks of raw materials for use in production

Include items such as:
- materials used in production
- containers and packaging
- livestock (bought for processing)

Opening stocks		Closing stocks	
$ 000	2140	$ 000	2240

40 Finished goods

Include items such as items manufactured or processed from stocks owned by this business (whether undertaken by this or any other contracted business).

Opening stocks		Closing stocks	
$ 000	2163	$ 000	2263

41 Work-in-progress

Include items such as:
- items being manufactured from stocks owned by this business (whether undertaken by this or any other contracted business) but not yet completed
- services provided by this business (whether undertaken by this or any other contracted business) but not yet completed
- work not billed at the end of the financial year

Opening stocks		Closing stocks	
$ 000	2164	$ 000	2264

42 TOTAL STOCKS
TOTAL of questions 38 to 41

Opening stocks		Closing stocks	
$ 000	2199	$ 000	2299
This is total C		This is total D	

Surplus or deficit calculation

43 Surplus or deficit before tax equals:

	Total income	from total A - page 4	$ 000	2301
minus	Total expenditure	from total B - page 7	$ 000	2302
minus	Opening stocks	from total C - page 8	$ 000	2303
plus	Closing stocks	from total D - page 8	$ 000	2304
	equals	SURPLUS / DEFICIT	$ 000	2390

If the calculated surplus or deficit is different from that shown in your annual accounts please give the reasons.

Use a minus sign to show a negative value.

Dividends

44 Total of dividends paid and dividends proposed by this business for the reported financial year

Dividends paid and proposed should be calculated:
- before the deduction of Resident Withholding Tax (RWT)
- before any allowance for imputation credits

Include:
- interim dividends
- proposed final dividends

Don't include final dividends from the previous year paid in this financial year.

$ 000 0013

Fixed assets

Include:
- all fixed assets as shown in your books of account
- all fixed assets operated by this business under finance lease arrangements
- leasehold improvements

Don't include:
- expenditure on maintenance
- intangible assets such as goodwill

Note:
- if any entries in the REVALUATIONS column are negative, or losses, please use a minus sign to show this, for example $ - 3 000 shows a loss of $3,000.
- entries in other columns should always be positive.

		OPENING BOOK VALUE at start of financial year	ADDITIONS (report at cost **Don't** include revaluations)	DISPOSALS (report at sale price **Don't** include devaluations)	REVALUATIONS include gain/loss on sale	DEPRECIATION (for reported financial year only)	CLOSING BOOK VALUE at end of financial year
45	Land	$ 000 (3501)	$ 000 (3502)	$ 000 (3503)	$ 000 (3504)		$ 000 (3506)
46	Buildings	$ 000 (3511)	$ 000 (3512)	$ 000 (3513)	$ 000 (3514)	$ 000 (3515)	$ 000 (3516)
47	Motor vehicles and other transport equipment	$ 000 (3521)	$ 000 (3522)	$ 000 (3523)	$ 000 (3524)	$ 000 (3525)	$ 000 (3526)
48	Computer software Include items such as packaged software and software developed by others for this business. **Don't** include software developed internally for this business. Include this as an intangible asset later.	$ 000 (3531)	$ 000 (3532)	$ 000 (3533)	$ 000 (3534)	$ 000 (3535)	$ 000 (3536)
49	Computer hardware Include items such as computers, monitors, printers, and multi-function devices. Note: if separate hardware and software figures are not available then give the combined amount here.	$ 000 (3541)	$ 000 (3542)	$ 000 (3543)	$ 000 (3544)	$ 000 (3545)	$ 000 (3546)
50	Lifting and handling equipment Include items such as: • forklifts • hoists • conveyors	$ 000 (3571)	$ 000 (3572)	$ 000 (3573)	$ 000 (3574)	$ 000 (3575)	$ 000 (3576)
51	Furniture and fittings	$ 000 (3551)	$ 000 (3552)	$ 000 (3553)	$ 000 (3554)	$ 000 (3555)	$ 000 (3556)
52	Other plant, machinery and equipment Include items such as: • machinery used for the production of goods • freezers	$ 000 (3631)	$ 000 (3632)	$ 000 (3633)	$ 000 (3634)	$ 000 (3635)	$ 000 (3636)
53	All other fixed assets Please describe the main items	$ 000 (3781)	$ 000 (3782)	$ 000 (3783)	$ 000 (3784)	$ 000 (3785)	$ 000 (3786)
54	TOTAL FIXED ASSETS Total of questions 45 to 53	$ 000 (3791)	$ 000 (3792)	$ 000 (3793)	$ 000 (3794)	$ 000 (3795)	$ 000 (3796) This is total E

Intangible assets

55 Computer software and / or databases developed internally by this business for its own use

Include items such as:
- client and customer databases developed internally
- software systems and programs developed internally

Don't include software already recorded in the fixed assets schedule on the previous page.

Leave answer boxes blank if this does not apply to this business.

opening book value (at start of financial year)	$ 000	3830
additions (report at cost; **don't** include revaluations)	$ 000	3831
disposals (report at sale price; **don't** include devaluations)	$ 000	3832
revaluations (include gain / loss on sale)	$ 000	3833
amortisation (for reported financial year only)	$ 000	3834
closing book value (at end of financial year)	$ 000	3835

56 Other intangible assets

Include items such as:
- goodwill, brands, trademarks
- rights to use a copyright, patent, software or knowledge produced from research
- fishing quota / ITQ

Don't include any computer software or database intangible assets already recorded in question 55

opening book value (at start of financial year)	$ 000	3820
additions (report at cost; **don't** include revaluations)	$ 000	3821
disposals (report at sale price; **don't** include devaluations)	$ 000	3822
revaluations (include gain / loss on sale)	$ 000	3823
amortisation (for reported financial year only)	$ 000	3824
closing book value (at end of financial year)	$ 000	3825

Please describe the main items

Capital work undertaken by own employees

57 In the last financial year has this business **capitalised** any salaries or wages?

- yes
- no → go to 61

2805

58 Estimate the amount of capitalised salaries or wages paid $ 000 2803

59 Estimate the amount of all other capitalised costs related to work undertaken by the business's **own** employees

Include items such as materials. $ 000 2804

60 Describe the main assets that were produced that relate to these capitalised costs.

Financial position

61 Please complete as at the balance date of the financial year covered by this questionnaire.
All amounts should be at closing book value.

Current assets

62 Cash and cash equivalents Include items such as bank deposits.	$	000	2569
63 Debtors Include items such as taxes receivable.	$	000	2571
64 Stocks (calculated in question 42, total D - page 8)	$	000	2572
65 Other current assets Include items such as: prepayments, short-term bills and bonds, shareholders' current accounts. Please describe the main items	$	000	2574

Fixed assets

66 Fixed assets (Book value calculated in question 54, total E - page 11)	$	000	2551

Non-current assets

67 Shares in associated and subsidiary companies	$	000	2583
68 Other shares	$	000	2584
69 Long-term bond and debenture investments	$	000	2585
70 Long-term loans Include items such as amounts owing by subsidiaries or group companies.	$	000	2586
71 Other non-current assets Include items such as intangibles, eg goodwill. Please describe the main items	$	000	2587

Total assets

72 TOTAL ASSETS Total of questions 62 to 71	$	000	2599

This is total F

Equity / Shareholders' funds

73 Issued and paid-up share capital	$	000	2526
74 Accumulated retained earnings	$	000	2527
75 All other reserves and long-term provisions	$	000	2528

Use a minus sign to show a negative value.

Current liabilities

76 Bank overdrafts	$	000	3900
Other short-term bank loans	$	000	3901
Creditors	$	000	2532
Provisions (eg provision for GST, PAYE, income tax)	$	000	2533
Short-term bills	$	000	3902
Short-term bonds	$	000	3903
Short-term related party debt (eg debt to a parent company)	$	000	3904
Other current liabilities Please describe the main items	$	000	3905

Non-current liabilities

77 Long-term bank debt	$	000	3910
Long-term bonds	$	000	3911
Long-term debentures	$	000	3912
Convertible securities	$	000	2540
Other long-term loans	$	000	3913
Long-term related party debt (eg debt to a parent company)	$	000	3914
Other non-current liabilities Please describe the main items	$	000	3915

Total equity and liabilities

78 TOTAL EQUITY and LIABILITIES Total of questions 73 to 77 Note: total G should equal total F on the previous page.	$	000	2549

This is total G

Other details

79 How long did it take you to complete this questionnaire?

Include:
- the time spent reading the instructions, working on the questions and obtaining information
- the time spent by all employees in collecting and providing this information

hrs mins 0006

80 Please make any comments that would help Statistics New Zealand interpret the information that you have given.

81 Who should we contact if we have any queries about this questionnaire?

Name

Position

Email

Phone Fax

I declare that this questionnaire has been completed to the best of my knowledge.

Signature Date

Day Month Year

For your own records, please take a photocopy of the completed form before returning it.

Thank you for your time and effort.

The main results of all our surveys are available at www.stats.govt.nz

Office use: A B C

A B C

2013

11. Appendix 2(b): IR10 form

Inland Revenue
Te Tari Taake

Accounts information

IR 10
November 2008

Year ended 31 March

- You only need to complete this form if you are in business. Please refer to the notes on page 3 – "About the IR 10".
- Please complete both pages of this form. Copy each amount from your financial statements.

Your full name

Your IRD number (8 digit numbers start in the second box. 1 2 3 4 5 6 7 8)

		Box	
Multiple activity indicator	Read Note 1 on page 3.	1	Yes / No
Gross income from	**Sales and/or services**	2	$ .00
Cost of goods sold	Opening stock (include work in progress)	3	$ .00
	Purchases	4	$ .00
	Closing stock (include work in progress)	5	$ .00
Gross profit	**(if a loss put a minus sign in the last box)**	6	$ .00
Other gross income	Interest received	7	$ .00
	Dividends	8	$ .00
	Rental and lease payments	9	$ .00
	Other income	10	$ .00
Total income	**(add up all income entered in Boxes 6 to 10—if a loss put a minus sign in the last box)**	11	$ .00
Expenses (do not include amounts treated as non-deductible for tax purposes)	Bad debts (written off)	12	$ .00
	Depreciation	13	$ .00
	Entertainment	14	$ .00
	Fringe benefit tax	15	$ .00
	Insurance (exclude ACC levies)	16	$ .00
	Interest expenses	17	$ .00
	Legal expenses	18	$ .00
	Rates	19	$ .00
	Rental and lease payments	20	$ .00
	Repairs and maintenance	21	$ .00
	Research and development	22	$ .00
	Salaries and wages	23	$ .00
	Subcontractor payments	24	$ .00
	Travel and accommodation	25	$ .00
	Vehicle (excluding depreciation)	26	$ .00
	Other expenses	27	$ .00
Total expenses	**(add up all expenses entered in Boxes 12 to 27)**	28	$ .00
	Total current year taxable profit (if a loss put a minus sign in the last box)	29	$ .00

Balance sheet items

		Box	
Current assets (as at balance date)	Accounts receivable (debtors)	30	$ .00
	Bank accounts	31	$ .00
	Other current assets	32	$ .00
	Total current assets	33	$ .00
Fixed assets (closing tax value)	Vehicles	34	$ .00
	Plant and machinery	35	$ .00
	Furniture and fittings	36	$ .00
	Land and buildings	37	$ .00
	Other fixed assets	38	$ .00
	Total fixed assets	39	$ .00
Other assets (as at balance date)	Intangibles	40	$ .00
	Preference shares	41	$ .00
	Shares and debentures	42	$ .00
	Term deposits	43	$ .00
	Other assets	44	$ .00
Total assets		45	$ .00
Current liabilities (as at balance date)	Accounts payable (creditors)	46	$ .00
	Bank accounts	47	$ .00
	Other current liabilities	48	$ .00
	Total current liabilities	49	$ .00
	Term liabilities	50	$ .00
Total liabilities		51	$ .00
Proprietor or shareholder funds	Drawings	52	$ .00
	Current accounts closing balance (if a **debit put a minus sign in the last box)**	53	$ .00
Total proprietor or shareholder funds	**(if a debit put a minus sign in the last box)**	54	$ .00
Other information	Deductible loss on disposal of fixed assets	55	$ .00
	Capital gain on disposal of fixed assets	56	$ .00
	Dividends paid	57	$ .00
	Are your accounts GST-exclusive?	58	Yes / No
	Are your accounts for a period of 12 months?	59	Yes / No

Cut off this page and staple it to page 3 of your return. You do not need to send in your financial records as well.

2

RESET form

Recent Motu Working Papers

All papers in the Motu Working Paper Series are available on our website www.motu.org.nz, or by contacting us on info@motu.org.nz or +64 4 939 4250.

15-14 Grimes, Arthur, Robert MacCulloch and Fraser McKay. 2015. “Indigenous Belief in a Just World: New Zealand Maori and other Ethnicities Compared.” (forthcoming)

15-13 Apatov, Eyal, Richard Fabling, Adam Jaffe, Michele Morris and Matt Thirkettle. 2015. “Agricultural Productivity in New Zealand: First estimates from the Longitudinal Business Database.”

15-12:Laws, Athene, Jason Gush, Victoria Larsen and Adam B Jaffe. 2015. “The effect of public funding on research output: The New Zealand Marsden Fund.” (forthcoming)

15-11 Dorner, Zachary and Suzi Kerr. 2015. “Methane and Metrics: From global climate policy to the NZ farm.”

15-10 Grimes, Arthur and Marc Reinhardt. 2015. “Relative Income and Subjective Wellbeing: Intra-national and Inter-national Comparisons by Settlement and Country Type”

15-09 Grimes, Arthur and Sean Hyland. 2015. “A New Cross-Country Measure of Material Wellbeing and Inequality: Methodology, Construction and Results.”

15-08 Jaffe, Adam and Trinh Le. 2015. “The impact of R&D subsidy of innovation: a study of New Zealand firms.”

15-07 Duhon, Madeline, Hugh McDonald and Suzi Kerr. 2015 “Nitrogen Trading in Lake Taupo: An Analysis and Evaluation of an Innovative Water Management Policy.

15-06 Allan, Corey, Suzi Kerr and Campbell Will. 2015. “Are we turning a brighter shade of green? The relationship between household characteristics and greenhouse gas emissions from consumption in New Zealand” (forthcoming)

15-05 Fabling, Richard and Lynda Sanderson. 2015. "Exchange rate fluctuations and the margins of exports”

15-04 Fabling, Richard, Richard Kneller and Lynda Sanderson. 2015. "The impact of tax changes on the short-run investment behaviour of New Zealand firms"

15-03 Sin, Isabelle, Steven Stillman. 2015. “Economic Liberalisation and the Mobility of Minority Groups: Evidence for Māori in New Zealand”

15-02 Grimes, Arthur, Ian Mitchell. 2015. “Impacts of Planning Rules, Regulations, Uncertainty and Delay on Residential Property Development”

15-01 De Rassenfosse, Gaétan, and Adam B. Jaffe. 2015. “Are Patent Fees Effective at Weeding Out Low-Quality Patents?”

14-15 Sin, Isabelle, Richard Fabling, Adam Jaffe, David C. Maré and Lynda Sanderson. 2014. “Exporting, Innovation and the Role of Immigrants.”

14-14 McLeod, Keith, Richard Fabling and David C. Maré. 2014. “Hiring New Ideas: International Migration and Firm Innovation in New Zealand.”

14-13 Timar, Levente, Arthur Grimes and Richard Fabling. 2014. “That Sinking Feeling: The Changing Price of Disaster Risk Following an Earthquake.”

14-12 Dorner, Zack, with Dean Hyslop. 2014. “Modelling Changing Rural Land Use in New Zealand 1997 to 2008 Using a Multinomial Logit Approach.”