

**Geographic concentration and firm
productivity**

David C. Maré & Jason Timmins

**Motu Working Paper 06–08
Motu Economic and Public Policy Research**

October 2006

Author contact details

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

Jason Timmins
New Zealand Department of Labour
PO Box 3705
Wellington
New Zealand
Email: jason.timmins@dol.govt.nz

Acknowledgements

The work contained in this paper was funded by the New Zealand Ministry of Economic Development (MED) and with funds from Motu's FRST-funded research programme on "Understanding Adjustment and Inequality". Thanks to the following people, who provided valuable comments on earlier drafts of the paper: William Strange, Helen Simpson, Ron Crawford, Richard Fabling, and Bettina Schaer. Jason Timmins' work on the project was done while he was an employee of Motu.

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the authors, not Statistics New Zealand, and do not necessarily reflect the views of the Ministry of Economic Development.

Motu Economic and Public Policy Research
PO Box 24390
Wellington
New Zealand

Email info@motu.org.nz
Telephone +64-4-939-4250
Website www.motu.org.nz

© 2006 Motu Economic and Public Policy Research Trust. All rights reserved. No portion of this paper may be reproduced without permission of the authors. Motu Working Papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review or editorial treatment. ISSN 1176-2667.

Abstract

Firms operating in dense labour markets are more productive, although understanding the mechanisms behind this relationship is both challenging and contentious. This paper uses a newly assembled dataset on location and labour productivity of most New Zealand firms to examine the role of location patterns at the industry, local labour market, and industry*location levels. We derive estimates in the presence of firm, location, and period fixed effects, paying particular attention to controlling for unobserved local and industry factors. Our findings confirm that labour productivity is higher for firms in geographically-concentrated industries (“localisation”), for firms in more industrially-diversified labour markets (“urbanisation”), and for firms operating in larger labour markets. Controlling for heterogeneity of industries, locations, and firms, we find some support for a positive productivity effect of changes in both localisation and urbanisation, although not all estimated effects are statistically and economically significant.

JEL classification

R12 - Size and Spatial Distributions of Regional Economic Activity;

R3 - Production Analysis and Firm Location

Keywords

Labour Productivity, Geographic concentration; agglomeration

Contents

1	Background.....	1
2	Approach	4
2.1	Regression model.....	4
3	Data.....	7
3.1	Business Demography Statistics (BD).....	7
3.2	Business Activity Indicator Series	12
3.3	Data subset.....	14
3.4	Measuring labour productivity.....	15
4	Analysis	17
4.1	Descriptive Statistics.....	17
4.2	Agglomeration and productivity - Regression analysis	23
4.3	Variation in agglomeration effects.....	30
5	Summary and conclusions	36
	Appendix A : Minimum industry coverage	41
	Appendix B : Statistics New Zealand Guide to interpreting the Business Activity Indicator data.....	42
	Appendix C Matching BD and BAI records (2003)	43
	Appendix D : Labour productivity means tables	44

Table of figures

Figure 1:	Map of Labour Market Areas.....	10
Figure 2:	Distribution of productivity - Enterprise and employment shares by productivity band (pooled across periods).....	18
Figure 3	Share of FTE employment by labour productivity and firm size bands (pooled across periods).....	20
Figure 4	Share of FTE employment by labour productivity and non-exporting and exporting firms (pooled across periods).....	21
Figure 5:	Productivity variation – by period industry, and LMA.....	19

Tables

Table 1: Prevalence of Imputation in matched sample – 1999/2000 to 2002/03 ..	13
Table 2 Sample selection.....	15
Table 3 Labour productivity for births, deaths and continuing firms.....	17
Table 4 Regression means and Standard deviations table.....	22
Table 5 Geographic concentration indices and firm characteristics separately.....	25
Table 6 Controlling for industry, LMA, and enterprise effects.....	26
Table 7: Economic significance of estimated effects	29
Table 8 Variation in the nature and strength of agglomeration effects – size, age, and exporting.....	32
Table 9 Variation in the nature and strength of agglomeration effects – Industry variation	34
Table 10: Colocation Groups.....	36
Appendix Table 2: Industry coverage	41
Appendix Table 3: Impact of Industry Coverage Restrictions	41
Appendix Table 1: Matching enterprises between the BD and BAI datasets in 2003	43
Appendix Table 4: Mean (under-represented) productivity by LMA [23/11] ..	44
Appendix Table 5: Mean (under-represented) productivity by Industry.....	44

Firms operating in dense urban areas are more productive. Ciccone and Hall (1996) find that doubling employment density in a U.S. county increases average labour productivity by 6 percent. Similarly, Henderson (2003) finds that a ten-fold increase in the number of local plants in a high-tech industry increases labour productivity by over 20 percent. Identifying the mechanisms that give rise to these relationships is, however, both challenging and contentious. In a recent broad review of evidence on the nature and sources of urban agglomeration economies, Rosenthal and Strange (2004) summarise debates about the industrial, geographic and temporal scope of agglomeration economies. They conclude that “there is a lot that we do not yet know about agglomeration economies”.

Rosenthal and Strange also note the importance to policy-makers of understanding agglomeration economies. Current New Zealand policy interest in ‘sustainable cities’, regional development, cluster development, and ‘Regional Centres of Excellence’ reflects an acceptance that local factors matter for firm performance. Knowing more about why and how they matter could contribute to better designed and more effective policies.

The current paper contributes to the literature by examining the relationship between geographic concentration and labour productivity for New Zealand firms over the period from 1993/4 to 2002/3. It considers a range of summary measures of geographic concentration, capturing differences between industries, between local labour markets, and between ‘city-industries’. In particular, it separately identifies the contributions of different agglomeration mechanisms to labour productivity, paying particular attention to controlling for area, industry, and firm heterogeneity. Furthermore, it examines variation in the strength of agglomeration effects across a range of observable dimensions such as firm size, industry, private versus public sector, and exporting status.

Our findings confirm that labour productivity is higher for firms in geographically-concentrated industries (“localisation”), for firms in more industrially-diversified labour markets (“urbanisation”), and for firms operating in larger labour markets. Controlling for heterogeneity of industries, locations, and firms, we find some support for a positive productivity effect of changes in both localisation and urbanisation, although not all estimated effects are statistically and economically significant.

1 Background

One of the main reasons that there is not general agreement about the nature and sources of agglomeration economies is that there are many theories that are consistent with the observation that cities have higher productivity. In a review of agglomeration theories, with a particular emphasis on cities, Duranton and Puga (2004) , summarise the challenge as follows:

“different microeconomic mechanisms may be used to justify the existence of cities. These mechanisms generate final outcomes that are observationally equivalent in most (but not all) respects. This ‘Marshallian’ equivalence is partly good news in the sense that the concept of urban agglomeration economies is robust to many

different specifications and microeconomic mechanisms. But this equivalence is also partly bad news because empirically identifying and separating these mechanisms becomes very difficult.” [p. 40]

The current paper quantifies the nature and impact of agglomeration effects, but does not attempt to identify the particular mechanisms that give rise to such effects. Even for studies that do not endeavour to identify the mechanisms of agglomeration, the evidence on the existence of agglomeration effects can take many forms. Rosenthal and Strange (2004) distinguish between direct and indirect approaches to estimation. Direct estimation relates to estimation of productivity effects by estimating a firm-level production function, augmented by measures capturing the firm’s environment:

$$y_j = g(A_j) f(x_j) \quad (1)$$

where y_j is the output of firm j , x_j are the inputs used by firm j , and A_j captures the impact of agglomeration, which can vary along three dimensions - industrial scope, geographic scope, and temporal scope.¹

The challenges of getting reliable measures of productivity have led some researchers to use what Rosenthal and Strange (2004) refer to as ‘indirect’ approaches to measurement. They cite studies that use employment growth, firm birth rates, wages, or rents, all of which could reasonably be expected to reflect or respond to productivity differences. Each approach has its own strengths as well as estimation challenges, which are well summarised by Rosenthal and Strange (2004).

As Rosenthal and Strange (2004) point out, many empirical studies of agglomeration focus on a particular dimension of the scope of agglomeration effects (industrial, geographical, temporal). One consequence of restricting attention to just one dimension of agglomeration effects is that agglomeration can, in principle, be detected using aggregated rather than firm-level data – using observations on locations, on industries, or on ‘city-industry’ observations. The decision to pursue estimation of agglomeration effects using variation in aggregated measures of firm performance is often a consequence of limited data availability – the absence of firm-level data that would yield direct estimates of productivity.

Ciccone and Hall (1996) examine the relationship between average labour productivity and employment density using data on US states,² without consideration of industrial scope. Ciccone (2002) performs a similar analysis for Europe, although for smaller geographic areas. Both studies find a positive relationship between employment density and productivity, with a doubling of local density raising productivity by around 6 percent.

The investigation of industrial scope is often confined to testing for two types of agglomeration effects - localisation and urbanisation. (Moomaw (1998))

¹ It is common to assume that the impact of agglomeration is Hicks-Neutral with respect to other factors, as is implied by the form of equation (1).

² Their density measure is calculated from county-level employment densities, and is aggregated to state-level.

Localisation effects arise where a firm benefits from congregating with other firms from the same industry, whereas the term urbanisation captures the benefits of industrial diversity. 'Urbanisation' can also refer to the effects of local market scale *per se*, although we describe these pure scale effects separately. Localisation effects are sometimes referred to as 'Marshallian', after Marshall (1920, p.271), who included a clear discussion of the phenomenon. The three main sources of Marshallian agglomeration effects are labour market pooling, knowledge spillovers, and input-output linkages.

Although Marshall also discussed the impact of urbanisation (Marshall (1920, p.273-4)), these effects are more commonly associated with Jacobs (1969), whose work emphasised the advantages of local industrial diversity. There is a long-standing debate about whether localisation or urbanisation is the more important feature explaining the existence and performance of cities.

The impact of either urbanisation or localisation can, in principle, be detected by relating firm performance to agglomeration indicators measured at the location, industry, or location*industry level. At the location level, the diversity of industry mix is positively related to average firm productivity in the presence of urbanisation effects. In the presence of localisation effects, this relationship is negative for a given location size, since industry diversity can be achieved only with a lower average degree of own-industry concentration and hence less scope for localisation effects. Similarly, holding scale constant, an industry-level measure of geographic concentration is positively related to productivity in the presence of localisation effects, and negatively related in the presence of urbanisation effects. Agglomeration indicators measured for location-industries combinations (often city-industries or state-industries) offer greater scope for identification of different agglomeration effects. Such indicators provide the most direct measures of local own-industry concentration. A variety of related measures have been used to summarise the concentration of industry i in location j , such as location share of industry employment (E_{ij}/E_i), industry share of local employment (E_{ij}/E_j), or locational quotient ($LQ = E_{ij}E/E_iE_j$). Each of these would be positively related to the average productivity of firms in the city-industry if localisation effects dominate.

Analysing patterns for city-industries also allows researchers to focus on particular industries, where agglomeration effects are expected to be most evident. Glaeser et al (1992) examine the average (employment growth) performance of the six historically dominant industries in each of 170 US cities. Own-industry concentration is measured as the percent of local employment accounted for by these six industries, relative to the industry's share of national employment. Local industrial (non-)diversity is measured as the percent of local employment accounted for by the five next-largest industries. Glaeser et al (1992) find stronger evidence for urbanisation than for localisation. Henderson (2003) focuses attention on 4 high-tech industries and 5 large machinery industries, using the number of own-industry plants as the measure of concentration. To capture urbanisation, Henderson uses an area-wide (non-)diversity index, or measures of area size. In contrast to the Glaeser et al (1992) findings, Henderson finds no evidence for urbanisation effects, and localisation effects for high tech industries only.

With agglomeration indicators being measured for industries, locations, or location-industries, estimation could proceed using observations at any of these levels of aggregation. The use of less aggregated data, however, allows for better control for firm or worker heterogeneity or endogenous selection of high-performing firms and workers into denser or more concentrated areas and industries. Henderson (2003) uses firm-level observations, and models a firm production function, augmented with agglomeration indicators. He is able to control for firm heterogeneity by using observable firm characteristics and with plant fixed effects, and tests the effects of instrumenting at the plant-level for the endogeneity of plant inputs and agglomeration variables. Using individual-level observations, Glaeser and Maré (2001) estimate the wage impact of agglomeration, and include various controls for observable and unobservable individual heterogeneity and selection.

2 Approach

The approach taken in the current paper is to estimate the relationship between firms' average labour productivity and variables that capture patterns of agglomeration. In particular, we include agglomeration indicators measured at three different levels of aggregation, for industries (i), for local labour markets (j), and for industry-locations (ij). Industry and location dimensions of agglomeration are treated symmetrically in our estimation. By using multiple observations on individual firms (k) in different time periods (t), we are also able to control for firm heterogeneity.

2.1 Regression model

We estimate the relationship between a firm's labour productivity and different measures of employment concentration, using OLS regression. The model is given by:

$$f_{ijkt} = \alpha + \beta_1 X_{it} + \beta_2 V_{jt} + \beta_3 Z_{ijt} + \beta_4 W_{kt} + e_{ijkt} \quad (2)$$

where i denotes industry, j denotes location, k denotes an enterprise and t denotes time period. Each observation is for an enterprise in an industry, in a local labour market area (LMA), in a year. All covariates are measured as at the beginning of the period for which the labour productivity levels are measured. Regressions are weighted by initial employment in the enterprise, to reduce the undue influence of some very small outliers.

The dependent variable f_{ijkt} is a measure of an enterprise's labour productivity (value added per full-time equivalent worker). The key agglomeration variables of interest are the locational quotient, to capture localisation effects, and a Maurel-Sédillot (Maurel and Sedillot (1999)) index of industrial specialisation at the LMA level, capturing urbanisation.

The locational quotient ($\ln LQ_{ijt}$) varies by industry-location-time, and is thus captured in the Z_{ijt} term of Equation (2). It is an index of LMA-industry size, normalised for the size of the LMA and the size of the industry, and thus measures the degree to which an industry is over-represented in a particular LMA.

A positive coefficient on this variable is consistent with localisation effects. The formula for the log of the locational quotient is:

$$\ln LQ_{ijt} = \ln E_{ijt} - \ln E_{it} - \ln E_{jt} + \ln E_t \quad (3)$$

The Maurel-Sédillot index of industrial specialisation at the LMA level (LMA_MS_{jt}) captures urbanisation effects. The measure varies by location*time. Urbanisation is associated with a positive impact of local industry diversity, and thus by a *negative* coefficient on LMA_MS_{jt} . We therefore multiply the variable by -1 so that a positive coefficient represents a positive effect of urbanisation. The index is similar to the index of geographic concentration presented in Maurel and Sedillot (1999) but with the roles of industry and area reversed. Specifically, the formula for a Maurel-Sédillot index of local industrial specialisation is:

$$LMA_MS_{jt} = \left(\frac{\sum_j s_{jt}^2 - \sum_j x_{jt}^2}{1 - \sum_j x_{jt}^2} - H_{jt} \right) / (1 - H_{jt}) \quad (4)$$

where s_{jt} is the year- t proportion of location j 's employment that is in industry i and x_{jt} is the year- t proportion of all employment that is in industry i . The index takes a value of 1 when all of a location's employment is from a single industry, and takes a value of zero when the distribution of s_{jt} is the same as that of x_{jt} .

Symmetrically, we include as an industry variable the degree of geographic specialisation, measured as a Maurel-Sédillot index (MS_{it}). The formula for this measure is similar to equation 4, but with the area (j) subscript replaced by industry (i). It is thus identical to the measure presented in Maurel and Sedillot (1999). In the presence of the $\ln LQ_{ijt}$ measure of localisation, the impact of MS_{it} reflects the industry-level advantages of geographic concentration beyond that experienced by firms that are in localised areas. Such an effect may arise if localisation improves overall industry functioning, which may arise through facilitation of industry-wide coordination. It is included in part to ensure the symmetric treatment of industry and area dimensions of agglomeration in the regression equation.

Both localisation and urbanisation are limited by scale. For purely statistical reasons, larger industries are likely to be less geographically concentrated, and larger areas are likely to have a more diverse range of industries. To allow for this, and also to allow for scale itself to be a distinct source of agglomeration effects, we include overall industry size ($\ln E_{it}$) and area size ($\ln E_{jt}$), each measured as the log of FTE employment.³

A small number of plants in an area mechanically raises the degree of industrial specialisation. In the limit, an area with a single plant by definition must have high industrial specialisation. The Maurel-Sédillot indices control for this mechanical relationship, as evidenced by the appearance of the H_{jt} term in equation 4. A small number of plants may, however, have an independent effect

³ A firm's own employment is not removed from these industry and area aggregates, to allow for the possibility of within-firm spillovers.

on measured labour productivity, through local monopsony effects or through monopoly power at the industry level. Plant level Herfindahl indices are thus included for both industry ($Herf_{it}$) and area (LMA_Herf_{it}). Less competitive industries are expected to have higher measured labour productivity, due to higher markups, and the coefficient on $Herf_{it}$ is thus expected to be negative. Firms in thin local labour markets are able to achieve lower labour costs, and thus possibly undercut product-market competitors, yielding a negative effect of LMA_Herf_{it} on measured labour productivity.

We control also for variation in observed characteristics of firms, as shown by W_{kt} in equation 2. Firm size is captured by 6 size group dummies, based on full-time-equivalent employment.⁴ The age of the firm is captured by dummy variables for 6 age groups, measured in years.⁵ We also include an indicator for whether an enterprise has exported good or services overseas.⁶ These firm-level controls are included to control for heterogeneity of firms across locations and industries.

Firm location may be endogenous, with more productive firms choosing to agglomerate more, biasing estimates of the impact of agglomeration. To the extent that endogenous locational choice is captured by the included observable characteristics (or firm fixed effects), the inclusion of these control variables will reduce the endogeneity bias. While this is clearly not an ideal solution to potential endogeneity, we note that Henderson (2003, p.18) tries alternative corrections for endogeneity in his study, and concludes that fixed effect controls are sufficient. Furthermore, we would expect endogeneity bias to lead to an overstatement of the impact of agglomeration. Given that we find only weak evidence of agglomeration effects, we have chosen not to proceed further with instrumental variables estimation.

Finally, e_{ijkt} is an error term, the structure of which varies across different specifications. Period fixed effects are included in all regressions. In addition, some regressions have industry fixed effects, location fixed effects, firm fixed effects, or a combination of these.

The interpretation of the agglomeration coefficients when location fixed effects are included is quite different from that in a model without such fixed effects. In a standard spatial equilibrium model such as Roback (1982), any productive local amenity (which would include many forms of positive agglomeration effect) would be reflected in higher wages and land rents. Workers would be indifferent between a high-wage/ high-rent location and one that offers lower wages with lower rents. In equilibrium, firms would locate in a high-

⁴ FTEs include working proprietors and employees.

⁵ Although the sample we use is restricted to 1992 and 2003 (firms that appear in both the BD and BAI datasets) it is possible to use the earlier years from the BD (1987-1991) to calculate the age of each enterprise up to a maximum age of 6 years. The firm-age variable is top-coded at 6 years to ensure consistency across the periods – it is not possible to distinguish ages over 6 years for firms observed in 1992, even though it is possible in later years.

⁶ We use the presence of zero-rated GST sales to identify enterprises that export. Zero-rated GST sales are primarily sales of exported good and services. However, the sale of a “going concern” business and fine metals are also zero-rated.

wage/high-rent location only if there is a productivity advantage in doing so. A ‘between-location’ correlation between productivity and agglomeration could thus reflect any cross-sectional difference in amenities, or any form of agglomeration.

In the presence of location fixed effects, identification is gained from variation across time within a location (and possibly also from variation across industries and firms unless industry and firm effects are also included). The benefits from time-invariant amenities, or from locational differences in the strength of agglomeration effects are absorbed by the location fixed effects. Coefficients from a location-fixed-effects specification thus indicate the impact on productivity of *changes* in the included agglomeration measures. The fixed effects estimates also reduce the impact of firm self-selection, since changes over time in agglomeration measures are largely exogenous to any individual firm, even though the choice of location may be endogenous.

3 Data

The data used in this study are taken from the Statistics New Zealand’s Business Demography (BD) datasets and the Business Activity Indicator (BAI) series. This paper provides the first documented attempt to merge the BD and BAI data.

The BD dataset provides annual longitudinal data on the majority of New Zealand businesses from 1987 to 2003, measured as at February each year. The BAI series provides monthly longitudinal data on all New Zealand businesses that have registered for GST⁷ from April 1992 to the present. Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975.

The data used are the most comprehensive available for New Zealand, and are thus the most representative of the New Zealand economy as a whole, despite the industry coverage and ‘economic significance’ restrictions outlined below.

3.1 Business Demography Statistics (BD)⁸

The target population for the BD datasets is ‘all New Zealand businesses’, although, as outlined below, there are some exclusions and variations over time in coverage. The business demography dataset is updated in February each year as an annual snap-shot from the Statistics New Zealand Business Frame at that point in time. From 1987 to 1994, the data are taken from the Statistics New Zealand Business Directory, and from 1994 to 2003, they are from the Statistics New Zealand Business Frame.

⁷. GST is New Zealand’s Goods and Services Tax, which is described more fully in section 3.2.

⁸ This section draws on Carroll et al (2002) and Statistics New Zealand (2004).

The data are collected from a combination of survey and administrative sources – primarily the Statistics New Zealand Annual Business Frame Update Survey (ABFU⁹) which has been conducted in mid-February each year, since 1987, and the Inland Revenue Department's (IRD) Client Registration File, which is the universe of GST registered enterprises.¹⁰

Data are available for business units (called activity units until 1996, and geographic units thereafter), and for enterprises. A business unit relates to a particular business site and an enterprise may contain several business units. Two sets of industry coding are available for each business unit. The primary industry code relates to the main activity of the business unit. Where a business unit provides ancillary services to other units in the same enterprise or group of enterprises, the ancillary industry code indicates the predominant activity of the units to which the services are provided. In this study, industry classification is based on the ancillary industry code, which is the classification that Statistics New Zealand uses for its published Business Demography analyses. In this paper, we focus on enterprises rather than geographic units, although we use information on geographic units (GUs) to identify the geographic distribution of enterprises. We will refer to enterprises as 'firms' for the remainder of this paper.

The criteria for including a GU in the data changed during the period of our study, although the following requirements were in force throughout:

- The GU was located in NZ; and
- The GU's primary and ancillary industry was in-scope, meaning that it was one of the target industries for the Business Frame, as outlined below; and
- The industry of the enterprise to which the GU belonged was in-scope.

In addition there were administrative rules in place that relate to the timing of information in the Statistics New Zealand database. Geographic units administratively 'birthed' during February of the year in question were excluded, GUs that ceased or were administratively "killed" during February of the year in question were included, and (until 1996) the GU had a data confirm date no later than 1 January of the year in question. All GST-registered enterprises recorded on the IRD's client registration file are continually monitored to determine if they meet the 'economic significance' criteria described below. In addition, non-employed enterprises are monitored using PAYE tax information to see if and

⁹ The ABFU survey is administered to all businesses except farm type agriculture enterprises, and those with no employees that are not part of a group of enterprises. Prior to 1997 the survey was called the Annual Business Directory Update Survey. The response rate to the ABFU survey is about 90% overall, but higher for larger firms. In the case of non-response, the BF carries forward the last known survey details. There are approximately 100,000 smaller enterprises, which are not covered by the ABFU. In addition, enterprises that indicate to the IRD that they have no paid employees have their data for working proprietors estimated from the data provided to the IRD. The ABFU collects a variety of information, including number of employees, overseas ownership and activities, location, and main activity.

¹⁰ The Client Registration File currently includes 530,000 enterprises. For GST-exempt financial services enterprises, Statistics New Zealand supplements the Client Registration File data using various sources, including association lists, financial reports, and a list of superannuation (pension) schemes from the Government Actuary. In addition, in order to ensure appropriate timing of firm births and deaths Statistics New Zealand uses a variety of other sources including its own surveys and media reports to identify businesses for entry onto and exit from the business frame.

when they begin to employ staff. When firms or GUs register for GST they are added (or 'birthed') onto the Business Frame, and are given a new reference number.¹¹ In practice, the selection criteria tend to be applied liberally, and the business frame continues to monitor a number of firms that fail to satisfy the criteria of economic significance. Where firms or GUs are sold, merged, or liquidated this will result in them de-registering for GST. A non-employing firm is removed from the business frame once it deregisters for GST or files 12 months of consecutive zero GST returns.

A major change in the data is the shift from GST-registration to economic significance, which occurred in 1994. From 1987 to 1994, GUs were included only if they belonged to a GST-registered firm (i.e. with GST sales of at least \$30,000). From 1994 the GU was included only if it belonged to an 'economically significant' firm, where a firm was regarded as economically significant if it met any one of the following criteria:

- Greater than \$30,000 annual GST expenses or sales;¹² or
- More than 2 full-time equivalent paid employees; or
- In a GST-exempt industry except residential property leasing and rental; or
- Part of a group of enterprises.

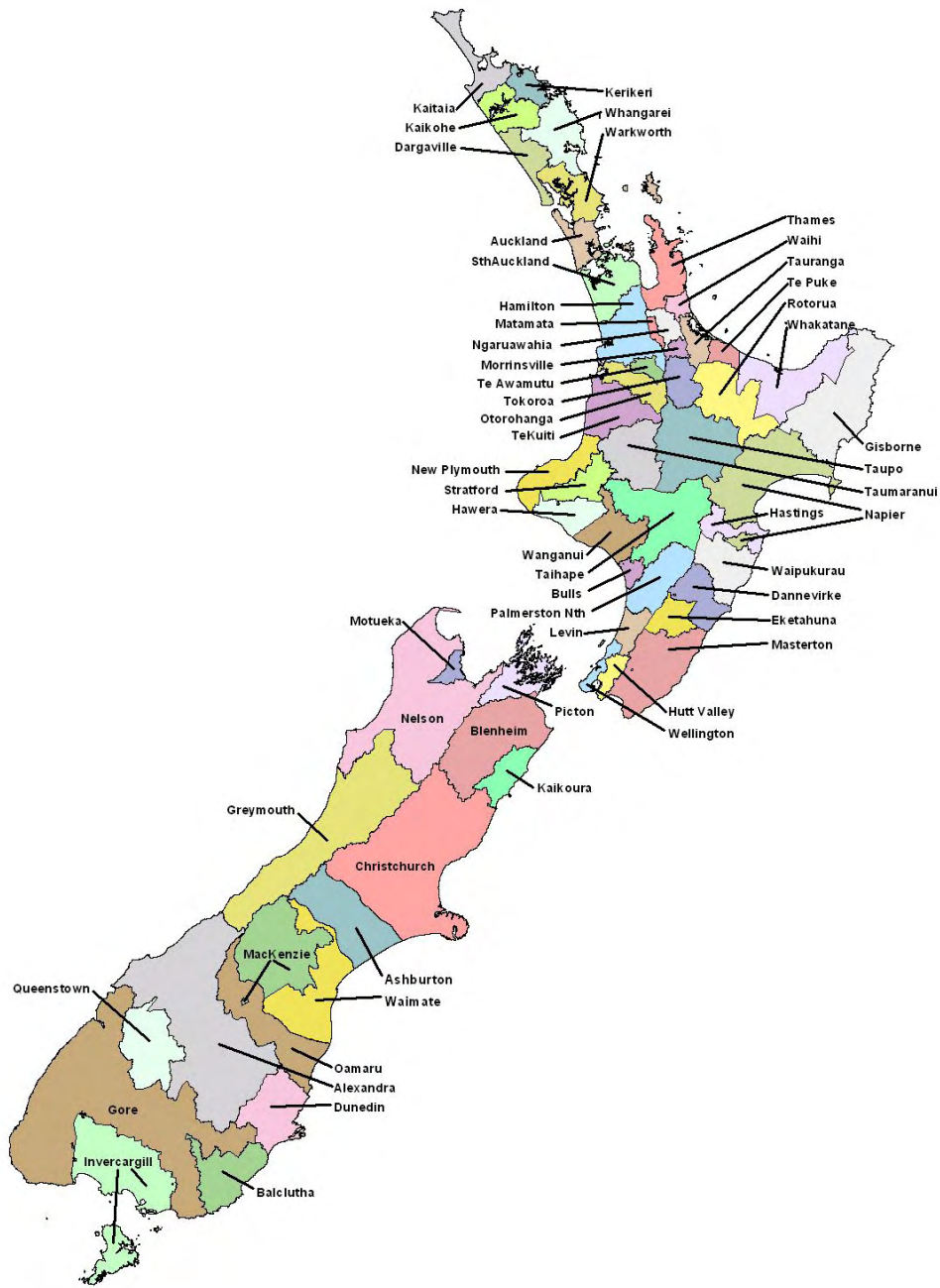
The economic significance definition thus excludes firms employing 2 or fewer FTE employees that were previously included, but adds in smaller firms that were in a GST exempt industry or were part of a group. The net effect was to decrease the number of firms. We measure employment in full-time equivalents, giving part-time employment half the weight of full-time employment. Working proprietors are included in this count of labour input.

In our analysis, we distinguish industries at the 4-digit level. The industry coverage of the business demography data has changed over time. The dataset does, however, provide a high degree of industry coverage throughout the period studied. In particular, it includes not only manufacturing industries, but also most service industries – a strength in comparison with datasets used in many related international studies.

¹¹ According to recent work carried out by Statistics New Zealand as part of the LEED (Linked Employer-Employee Data) initiative, "Births on the [Business Frame] that later turn out to be changes of ownership of geographic units already on the frame average approximately 15 percent of enterprise births per month" Seyb (2003), p. 14.

¹² The annual GST limit was set at \$30,000 from 1994, and increased to \$40,000 in October 2000 (IRD: GST Guide – November 2000). In practice Statistics New Zealand uses a GST 'buffer zone' of \$35,000 – \$45,000 in order to limit the extent of movements in- and out-of the BF because of the \$40,000 GST criteria: GST sales must exceed \$45,000 before being included, and fall below \$35,000 before being dropped. From 2001, enterprises were also included if their GST registration was compulsory, special or forced, which means that the business is expected to exceed the \$30,000 boundary. In 1994 enterprises satisfying both criteria were included, enabling a comparison of the sample frames for this year.

Figure 1: Map of Labour Market Areas



Notes: Labour Market Areas derived by Newell and Papps (2001).

The most notable exclusion from the BD is firms in agricultural production industries.¹³ Appendix Table 1 summarises the changing industry coverage restrictions throughout the period of the study. Until 1996 the industry selection criteria were based on the New Zealand Standard Industrial Classification (NZSIC); while from 1997 onwards the Australian and New Zealand SIC (ANZSIC) was used. We rely on ANZSIC codes for our analysis, even though these are derived from NZSIC codes prior to 1997, and accepting the caveat in Statistics New Zealand (2004) that the quality of industry coding will therefore be poorer in earlier years. Our central analyses in the paper will be restricted to industries that are within coverage throughout the 1987 to 2003 period. Only in 1998, when industry coverage was expanded to include agriculture, does this restriction cause a loss of more than 5 percent of (full-time equivalent) total employment. Appendix Table 2 summarises the impact on employment in each year of restricting industry coverage to continually covered industries.

The locations of firms are recorded at the level of meshblocks, which provide a very disaggregated level of geographical detail. Meshblocks range in size from city blocks to large areas of rural land. For many of our analyses, we look at the distribution of employment across larger areas, which are obtained by aggregating meshblocks. Our main analyses group employment into 58 labour market areas (LMAs), as defined by Newell and Papps (2001) on the basis of commuting patterns.¹⁴ Figure 1 provides a map of LMAs. A little over half of all employment is accounted for by the four largest LMAs, which contain the three largest metropolitan areas. In 2003, Auckland and South Auckland, in the upper North Island, accounted for around 21 percent and 14 percent of employment respectively. Christchurch, on the East coast of the South Island, accounted for 11 percent, and Wellington, the capital city, located at the bottom of the North Island, accounted for 9 percent. Of the remaining LMAs, the two largest abut metropolitan areas. Hamilton, just south of South Auckland, and Hutt Valley, East of Wellington, account for 5 percent and 3 percent of employment respectively. There are 12 LMAs that contain smaller centres, each accounting for between 1 and 3 percent of employment (Dunedin, Tauranga, Palmerston Nth, Nelson, Invercargill, Rotorua, Hastings, New Plymouth, Whangarei, Napier, Waimate, Wanganui) and the remaining 40 LMAs each account for less than 1 percent of total employment.

¹³ Between 1994 and 2001, the excluded industries were as follows: Agriculture and livestock production (NZSIC 11111-11199 in 1994-96; ANZSIC 01110-01699 in 1997, 1999-2001); Residential property leasing and rental (NZSIC 83121), Commercial property and leasing (NZSIC 83123), Child care services (NZSIC 93402), Residential and non-residential services (NZSIC 93403), and Business, professional and labour organisations (NZSIC 93500) in 1994-95; and Religious organisations (NZSIC 93910), Social and community groups (NZSIC 93990), and Sporting and recreational services (NZSIC 94402) in 1994-96.

¹⁴ Newell and Papps (2001) define two sets of labour market areas – one with 140 areas and one with 58. The main differences are that the 140-area set provides greater disaggregation of some relatively small areas. We have chosen to use the more aggregated areas because of the small size of some of the additional splits.

3.2 Business Activity Indicator Series

The Business Activity Indicator (BAI) Series includes all New Zealand businesses that have registered for Goods and Services Tax (GST). GST is a tax on the consumption of most goods and services in New Zealand. GST was introduced in New Zealand on 1 October 1986 at a rate of 10%, was increased to 12.5% in 1989, and is collected by the Inland Revenue Department (IRD). Prior to 1 October 2000 all businesses that conduct taxable activity were required to register for GST if their annual turnover was greater than \$30,000. From 1 October 2000, this threshold was increased to \$40,000.¹⁵

A few goods and services are GST exempt or zero-rated. GST exempt supplies include residential dwelling rentals, financial services and donated goods and services sold by non-profit organisations (e.g. club subscriptions).¹⁶ 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.¹⁷ However, the sale of an entire business (as a “going concern”), the first sale of fine metals by a refiner, certain financial services and imported supplies (supplied within New Zealand) are also zero-rated.

An enterprise is responsible for reporting GST and includes all taxable activities conducted by any activity units within the enterprise. However, within a group of enterprises it is possible for individual enterprises to report GST on behalf of other enterprises within the same group. Group reporting of GST results in missing GST information for the non-reporting enterprises. A further complication arises if the enterprise responsible for reporting GST within a group changes. Finally, an enterprise can choose to file their GST returns every month, bi-monthly, 6-monthly or 12-monthly. Group reporting of GST and different filing frequencies produce a raw GST file that contains missing GST data for some enterprises within a group and different GST filing frequencies.

The Business Activity Indicator (BAI) series is a derived monthly register of GST sales and purchases for all GST registered enterprises between April 1992 to present. The BAI series includes GST inclusive and zero-rated sales and GST inclusive purchases for each GST registered enterprise. To achieve a monthly record of GST returns for all GST registered enterprises it is necessary to reallocate GST sales and purchases for enterprises that do not report GST at

¹⁵ Any company, club or individual can register for GST even if their expected turnover is less than \$40,000.

¹⁶ A landlord will still pay GST on dwelling expenses, such as maintenance, rates and insurance. If a residential dwelling is sold as part of a taxable activity, the sale is GST exempt. Financial services include the following: paying or collecting any amount of interest, mortgages and other loans, bank fees, securities such as stocks and shares, providing credit under a credit contract, exchanging currency, arranging or agreeing to do any of the above, financial options, deliverable future contracts, non-deliverable future contracts. Any financial planning or monitoring services are subject to GST. Private non-profit organisations may still register for GST to enable them to claim GST credits from purchased goods and services.

¹⁷ Exported supplies include: exported goods, goods not in New Zealand at the time of supply, duty-free goods, exported vessels (ships), exported aircraft, goods and services that are directly connected with temporary imports, transport of passengers and goods to and from New Zealand, services performed outside New Zealand and certain exported services.

monthly intervals (2,6 and 12-monthly filers). The BAI dataset also imputes GST for an enterprise that does not file GST, or files in month when not expected to do so.¹⁸ To ensure that all enterprises have GST information the BAI series allocated GST returns across a group of enterprises using employment information from the BD and financial information collected by IRD using the IR10 form.¹⁹

Imputation was minimal prior to 1999/2000, with less than 0.5 percent of enterprises having one or more months of BAI data imputed during the year. Those enterprises also accounted for less than 0.5 percent of FTE employment in any year. Table 1 summarises the patterns after 1999/2000, when imputation rates increased. The first column lists annual periods, ending in February. Observations are restricted to enterprises that have at least one month of BAI data during the year, and which have BD information from either the start or end of the year. The share of observations with imputed GST information increases to 16 percent of enterprises, and 31 percent of FTE employment by 2002/03.²⁰ Given that, in most cases, imputation is made for only a single month of each firm's returns, and is based on information from the same firm's previous returns, observations with imputed information are retained in the sample.

Table 1: Prevalence of Imputation in matched sample – 1999/2000 to 2002/03

Period	Number of Enterprises (000)	Enterprises imputed (000)	%	Number of FTE (000)	FTE imputed (000)	%
1999/2000	303	14	4%	1,361	32	2%
2000/2001	312	39	13%	1,386	114	8%
2001/2002	313	47	15%	1,423	131	9%
2002/2003	323	53	16%	1,473	454	31%

Notes: All numbers have been randomly rounded. The analysis included only enterprise-year observations that have corresponding BD and BAI records. Periods refer to years ending in February.

3.2.1 Net GST sales as a proxy for value added

The GST data from the BAI are administrative data collected in the course of administering New Zealand's Goods and Services tax. As such, they are not designed to provide a measure of a firm's value added. They do, however,

¹⁸ An enterprise may file bi-monthly on odd months (e.g. February, April ... December), but happens to file on an even month. Further details of imputation methods is available at [http://www2.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Activity+\(GST\)+Indicator](http://www2.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Activity+(GST)+Indicator).

¹⁹ All businesses (that undertake an activity for profit) are required to fill in an IR10 form. The IR10 form collects information about a firm's income, expenses, assets and liabilities. See www.ird.govt.nz for more information.

²⁰ Analysing the monthly BAI series for 1999 to 2003, we find that the greatest share of imputations (81 percent) use information from the same firm's prior returns, either for firms that failed to provide a GST return when expected to (76 percent), or for firms that were not expected to file a return in a particular month (4 percent). Imputations for firms that did not file when expected to are made for at most one month of data. Growth in this type of imputation also accounts for most of the rise in imputation between 1998/99 and 2002/03. A further 20 percent of imputations are imputations made for the same reason, but where the necessary firm-specific information is unavailable. In these cases, imputation is based on mean monthly returns for firms in the same industry with the same filing frequency. Mean imputation may introduce significant month to month variation in the monthly series for a particular firm, although the impact on the annual GST information used in this paper will be less.

provide a useful proxy for firm productivity, and one that avoids some of the problems encountered in similar international studies.

In particular, they provide a very broad measure of the value of purchased goods and services, given that there are few exemptions to New Zealand's GST. The data contain information on purchases of all inputs, in contrast to the US county-level value-added data criticised by Ciccone and Hall (1996) because they do not include purchased services. Another advantage of the data is that they cover most types of economic activity, with the notable exception of agricultural production.

Value added is defined as the value of gross output, less the value of intermediate consumption. Appendix B contains information on the main differences between the BAI net sales series and value added. The main discrepancies are that the BAI measure includes sales and purchases of capital goods, sales and purchases of businesses, and excludes a small range of services as outlined above. The inclusion of capital purchases will tend to bias downwards estimates of productivity in years when purchases are made, and bias them upwards in other years. Over time, or across firms, these timing effects are expected to average out to some degree.

3.3 Data subset

The dataset used for analysis covers the period February 1993 to February 2003. The 1993/94 annual period is the first period where GST is collected in every month. We use a subset of these data. Every enterprise-year observation in our sample must have matching BAI and BD data. Furthermore, we limit attention to enterprises in industries that were within BD industry coverage from 1987/88 to 2002/03, and to observations that have non-missing information on employment, GST sales and purchases, industry, and LMA, and non-zero employment.²¹

Table 2 outlines the impact of our sample selection criteria on the number of enterprises and FTE employment. There are a total of 5,390,000 distinct enterprise-year combinations represented in the pooled BD and BAI data over the 1993/94 to 2002/03 period. Restricting attention to enterprise-years for which there is both a BD and a BAI record substantially reduces the number of observations.²² There are only 2,902,000 matched enterprise-year observations, which is 54% of the initial total. Most of the unmatched records are BAI records for which there is no matching BD record. Some such mismatch is to be expected, given that the BD data is for a restricted subset of enterprises – those that are economically significant, and that belong to the subset of industries that are covered by the BD.²³ In contrast, there are relatively few records in the BD that do not have a matching BAI record – presumably at least some of these are in

²¹ Data on firm age is derived from pre-1993 BD records, so consistent industry coverage is required back to 1987/88.

²² Appendix C provides a more detailed account of the match between BD and BAI data.

²³ For the 2002/03 year, 52 percent of unmatched BAI records had GST sales and purchases below \$40,000, compared with only 3 percent for matched records. Appendix C contains further detail of the characteristics of matched and unmatched firms for the 2002/03 year.

GST-exempt industries.²⁴ Only about 3 percent of the FTE employment that is recorded in the BD is dropped as a result of requiring a BAI match.

Table 2 Sample selection

Selection Criteria	Enterprises			Employment		
	(000)	% of Total	% of Matched	(000)	% of Total	% of Matched
Total	5,390	(100%)		13,678	(100%)	
Matched BD+BAI records	2,902	(54%)	(100%)	13,277	(97%)	(100%)
Valid data	2,230	(41%)	(77%)	12,356	(90%)	(93%)
Single LMA operation	1,591	(30%)	(55%)	6,048	(44%)	(46%)

Notes: Column percentages are in parenthesis. Graduated random rounding has been applied to the enterprise and FTE counts to protect confidentiality.

The ‘valid data’ row of Table 2 shows the number of enterprises and employment for the matched subset. Of this subset, we require enterprises to be within BD industry coverage,²⁵ to have positive employment and to have non-missing GST, industry and location information. These requirements further reduce the sample to 2,230,000 enterprises and 12,356,000 FTE employment, which accounts for 77% of the matched enterprises and 93% of the employment in the matched sample. It is this ‘valid data’ subset that is used for most of the analysis that follows.

The final row of Table 2 reports on the subset of enterprises that operate in only a single LMA (“single-location enterprises”, or SLEs), and that do not change industry. Location information is available for GUs rather than for enterprises. If an enterprise contains more than one GU, it may operate in more than one LMA (“multi-location enterprises”, or MLEs), and in more than one industry. In this case, LMA-level agglomeration measures need to be averaged over all the LMAs in which the enterprise operates. We use a weighted average, with LMA measures weighted by the share of enterprise employment that is in each LMA. For firms that operate in multiple industries, we assign the single industry code of the enterprise, and include a dummy variable for multi-industry firms in any regressions. The SLE sample is very restrictive, accounting for only 55 percent of matched records, and 45 percent of matched employment. We use the SLE sample to test robustness of our results, but not as a primary analysis sample.

3.4 Measuring labour productivity

The BD and BAI datasets are used to create a measure of firm level labour productivity. The BD dataset provides an annual snapshot (mid-February) of the level of employment within each enterprise and the BAI datasets provides a measure of a firm’s value added output (*netsales*) at a monthly interval.

Value-added output is proxied by a firm’s annual net sales. Annual net sales for firm k in year t is generated by summing net sales for each month (m) for which sales or purchase information is available.

²⁴ It is not possible to confirm this because the BAI data do not contain any industry identifiers.

²⁵ All information on a firm is dropped if the firm is out of coverage in any year.

$$ann_net_{kt} = \sum_{m=1}^{12} (sales_{ktm} - purchases_{ktm}) \quad (5)$$

Sales and purchases are separately deflated. Sales are deflated by Statistics New Zealand's Producer Price Index – Outputs, for the one-digit industry to which the firm belongs. Purchases are deflated by the Producer Price Index – Inputs, also by one-digit industry. The resulting value added is expressed in 2003 dollars.

To create a measure of average labour productivity, this output measure is divided by average labour input during the year. Employment is observed only once every twelve months. For firms that are in existence throughout the year, we have a measure of employment at both the beginning and end of the period, and average labour input of firm- k in year t is calculated as $\bar{E}_{it} = 0.5 * (E_{it-1} + E_{it})$.

For firms that operate for only part of a year, care is needed in estimating an appropriate measure of employment. There are three types of part-year firms - firms that enter or exit production (firm births and deaths), and firms that are 'seasonal' producers, in the sense that *netsales* is missing for some months. In each case, we construct two measures of labour productivity – one that understates productivity and one that overstates it.²⁶

For firms that are born during the period, we observe only end-of-period employment. An understated measure of productivity is based on the assumption that employment has been at this level throughout the firm's first year ($\bar{E}_{it} = E_{it}$). Alternatively, we could assume that the firm started with zero employment and grew steadily during its first year, with mean employment of $\bar{E}'_{it} = E_{it}/2$. This generates estimated labour input that is only half as big, and consequently a measure of labour productivity that is twice as high. We refer to this as an 'overstated' measure of productivity. Symmetrically, for firms in their final year of operation, we observe only initial employment, and we could assume labour input of $\bar{E}_{it} = E_{it-1}$ for an understated productivity measure, or of $\bar{E}'_{it} = E_{it-1}/2$ for an overstated measure.

For firms that do not report net sales activity in every month, we could assume either that the firm employed its workers throughout the year, or that workers were employed only in those months when net sales is reported. Assuming that employment is continuous generates high estimated employment and therefore low ('understated') estimated productivity since it is assumed that the output reported for fewer than twelve months was generated by maintaining average employment of $\bar{E}_{it} = 0.5 * (E_{it-1} + E_{it})$ throughout the year. Alternatively, if workers are employed only part of the year, the appropriate measure of labour

²⁶ In practice, actual productivity may be greater than the overstated estimate or less than the understated estimate. The terms are intended to indicate a relatively high and a relatively low measure. On average, productivity estimates using the overstated measure appear higher than would be expected and estimates using the understated measure appear too low.

input is $\bar{E}'_i = \bar{E}_i * (M_i/12)$ where M_i is the number of months for which netsales is observed.

Overall, the difference between the two labour productivity measures will depend on the number of entering and exiting firms and on how many continuing firms are seasonal producers. Mean productivity using the overstated measure is around 10% higher than the mean of the understated measure. Table 3 shows the FTE-weighted means and standard deviations of the two measures. Firms in their first year of existence have productivity that is only between 21 percent and 41 percent of that in continuing firms, depending on which measure is used. Dying firms also have relatively low productivity – between 43 percent and 85 percent, depending on the measure.

Table 3 Labour productivity for births, deaths and continuing firms

	Over-estimated labour productivity		Under-estimated labour productivity		Percent of Enterprises	Percent of FTEs
	Mean	Std dev	Mean	Std dev		
Births	\$24.7	(\$784.4)	\$12.3	(\$392.8)	15.2%	3.6%
Continuing firms	\$59.8	(\$364.1)	\$59.2	(\$344.9)	71.8%	92.3%
Deaths	\$50.6	(\$532.3)	\$25.3	(\$266.0)	13.0%	4.1%
Total	\$58.1	(\$395.5)	\$56.1	(\$344.0)	100.0%	100.0%

Notes: Mean productivity has been weighted by FTE employment. Firm exit and entries include rebirths (firms that re-enter the dataset). Underlying enterprise and FTE counts are based on randomly rounded counts.

In our regressions, we include separate dummy variables for entering firms, for exiting firms, and for seasonal firms. The coefficients on these dummy variables are negative when using understated productivity and positive when using overstated productivity. Apart from this intercept-effect, regression results are not significantly affected by the choice of measure. Unless otherwise stated, understated productivity is used for the remainder of the paper.

4 Analysis

For analysis purposes, we draw a 65 percent sample of enterprises, with each enterprise having equal probability of selection, regardless of length of time in the data or employment size.²⁷ The resulting analysis dataset contains 1,450,036 enterprise-year observations, covering 7,826,921 FTE-years of employment. All agglomeration measures are, however calculated using the entire sample.

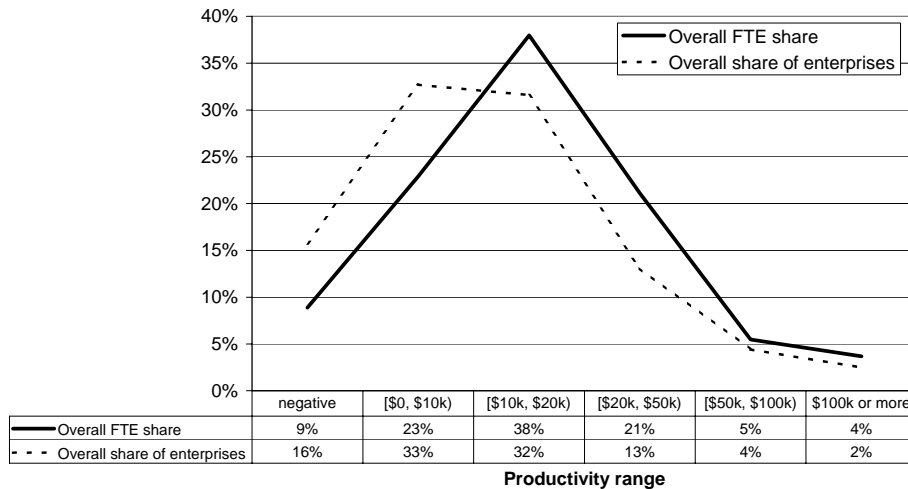
4.1 Descriptive Statistics

Figure 2 shows the distribution of average labour productivity across broad productivity bands. The dashed line show the proportion of enterprises in each band, and the solid line shows the proportion of FTE employment in each band. Sixteen percent of enterprises have negative productivity, although these tend to be relatively small enterprises, and account for 9 percent of employment.

²⁷ The decision to work with a sample rather than the full population is solely a consequence of computing resource limitations.

Overall, firm size and productivity levels are positively correlated. Just over half of enterprises have average labour productivity above \$20,000, and these enterprises account for just over two-thirds of employment. The average firm in the sample has labour productivity of \$35,900 whereas the average worker works in a firm with a labour productivity level of \$56,100, reflecting the fact that larger firms tend to have higher productivity.

Figure 2: Distribution of productivity - Enterprise and employment shares by productivity band (pooled across periods)



Notes: The productivity measure is ‘understated’ productivity – see text for details. The figure is based on a 65 percent sample of data. Underlying enterprise and firm counts have been randomly rounded.

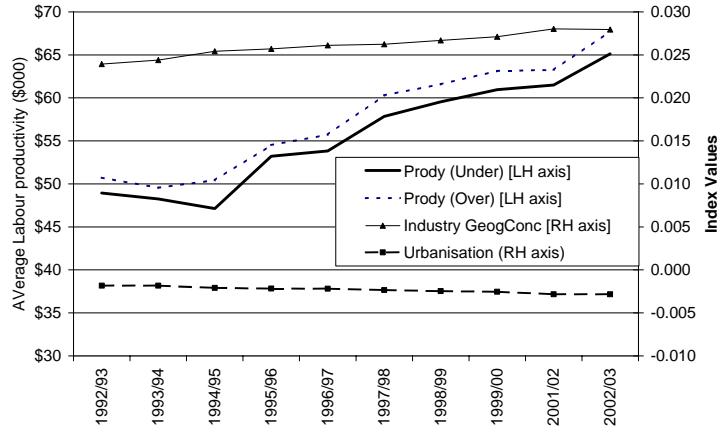
Productivity is observed at a firm level, whereas our measures of agglomeration are measured at industry, LMA, or Industry*LMA level. Figure 3 summarises the relationship between productivity and selected concentration variables across time, across industries, and across Labour Market Areas. The multivariate analyses that follow progressively control for these sources of variation to test further the link between location and productivity. Appendix D contains a tabular summary of mean productivity levels by LMA and by one digit industry.

Panel (a) of Figure 3 traces the growth over time in the two measures of productivity.²⁸ As noted above, we use an under-stated measure of average labour productivity in most of our subsequent analyses. The alternative (over-stated) measure closely tracks our chosen measure over time, with an average difference of around \$2,000. Productivity was trending upwards during our study period, with growth of around 30 percent over 10 years. There was also growth in average industry geographic concentration, as measured by the MS index, of 17 percent. Urbanisation decreased from -0.18% to -0.28%.

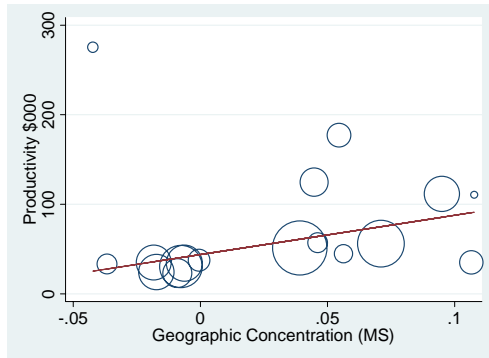
²⁸ The labour productivity growth of 33% over 9 years, as shown in Figure 3, is broadly similar to the 29% growth in Statistics New Zealand’s labour productivity series (Statistics New Zealand (2006)).

Figure 3: Productivity variation – by period industry, and LMA

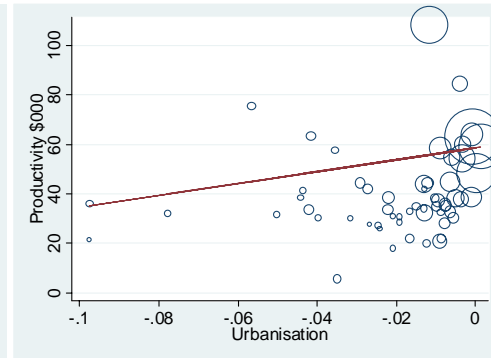
(a) By period



(b) By 1-digit industry



(c) By LMA



Notes: The productivity measure is ‘understated’ productivity – see text for details. The figures are based on a 65 percent sample of data. Underlying enterprise and firm counts have been randomly rounded. Panel (b) contains data on 17 1-digit industries. Panel (c) contains data on 58 LMAs.

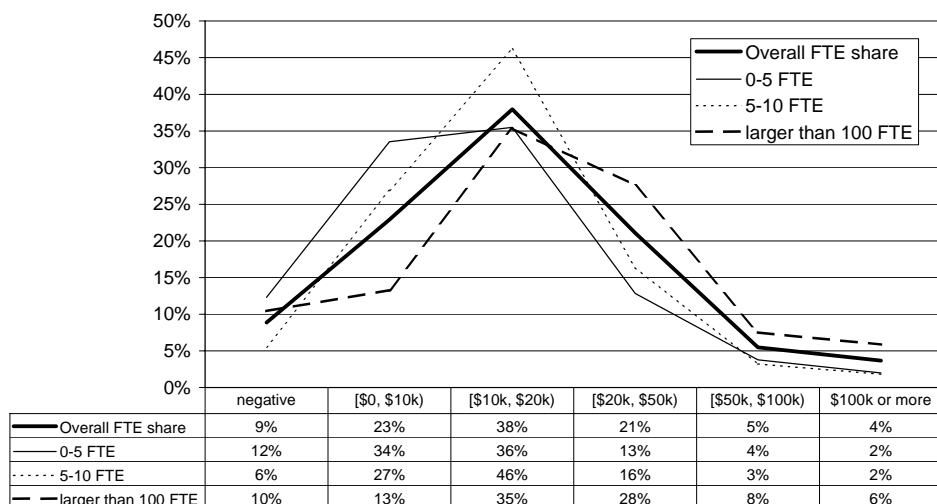
The second panel of Figure 3 shows the relationship at the level of industry. Each observation represents a one-digit industry, with geographic concentration and productivity both averaged over time. The graph symbols are scaled to reflect relative size in FTEs. There is a positive relationship between industry concentration and industry productivity, as shown by the positively sloped regression line.

Panel (c) shows the relationship between LMA urbanisation and mean productivity, with each observation representing a LMA. There appears to be a positive relationship, suggesting that urbanisation is associated with higher productivity, although the positively sloped regression line shown on the graph is not significantly different from zero.

Figure 4 shows the relationship between productivity and firm size in more detail. The solid line again shows overall FTE shares by productivity band.

The distribution for small firms (0 to 5 FTE) is shown by the lighter solid line. About a third of small firm workers are in firms with productivity of between \$0 and \$10,000, compared with only 23 percent of all employment. In contrast, 42 percent of workers in large firms (more than 100 FTE) are in firms with average productivity of more than \$30,000, compared with only 19 percent of small firm workers, and 30 percent overall.

Figure 4 Share of FTE employment by labour productivity and firm size bands (pooled across periods)

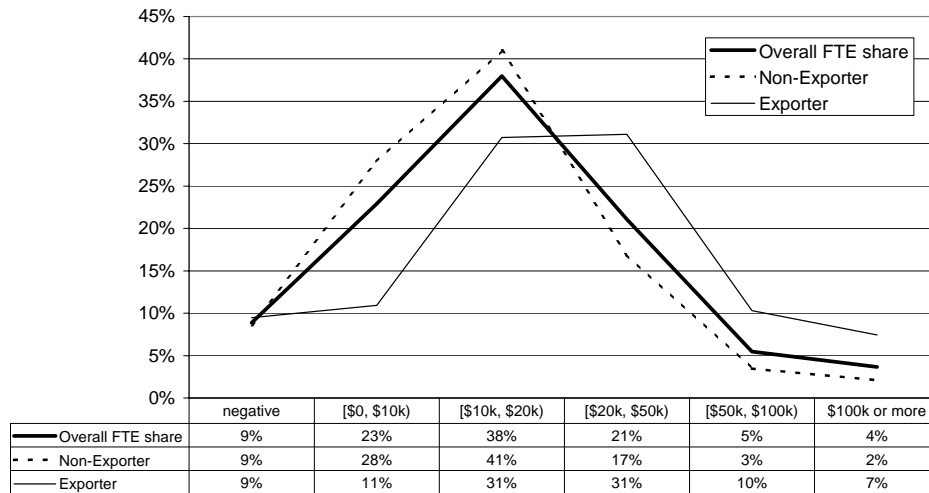


Notes: The productivity measure is 'understated' productivity – see text for details. The figure is based on a 65 percent sample of data. Underlying enterprise and firm counts have been randomly rounded.

Firm size is positively correlated with a firm's likelihood of exporting. Overall, 9.6 percent of enterprises report some export activity, and they account for 28.8 percent of FTE employment. As shown in Figure 5, exporting firms tend to have relatively high productivity, with 48 percent of workers in exporting firms in firms with productivity over \$20,000, compared with 22 percent of those in non-exporting firms.

Table 4 provides descriptive statistics for the variables that are used in the subsequent multivariate analysis of the relationship between productivity and firm location. All statistics are weighted by FTE employment. The first two columns provide sample means and standard deviations. The remaining columns show how the variation in each variable changes as we control for different sources of variation. These columns report 'within-group' standard deviations, for various group definitions, to provide an indication of the sources of variation that are used to identify fixed-effects regression coefficients.

Figure 5 Share of FTE employment by labour productivity and non-exporting and exporting firms (pooled across periods)



Notes: The productivity measure is 'understated' productivity – see text for details. The figure is based on a 65 percent sample of data. Underlying enterprise and firm counts have been randomly rounded,

Productivity measures are skewed to the right, and have high variance. Our chosen productivity measure has a mean of \$56,100 and a standard deviation of \$344,000. The overstated measure has a higher mean and standard deviation. Mean log industry size is 8.78, giving a geometric mean of 6,500 FTE. Similarly, the log size of the average LMA is 10.98, which is a geometric mean of 58,700. Exporters account for 28.8 percent of FTE employment. Recall that the measurement of productivity is problematic for births, deaths, and seasonal firms. On average, around 4 percent of FTE employment is in each of these categories. The average worker is in a firm with just under 600 workers, and one that has been in existence for well over 5 years.²⁹ By definition, larger firms employ a disproportionately large share of workers, as do older firms. If we do not weight firms by their FTE employment size, we find that the average firm employs 5.9 FTE workers, and has a mean age of 4.0 years.

²⁹ Note that firm age is top-coded at 6 years, so the reported average of 5.23 is an understatement of actual age. In regressions, age and size are both entered as sets of categorical variables.

Table 4 Regression means and Standard deviations table

Variable	Mean	SD	(a) within period	(b) within Industry	(c) within LMA	(d) LMA, Ind. & Yr FE	(e) Firm, LMA, Yr
Labour Productivity (\$000) (Underestimated)	\$56.1	\$344.0	\$344.0	\$332.5	\$343.2	\$332.1	\$279.8
Labour Productivity (\$000) (Overestimated)	\$58.1	\$395.5	\$395.5	\$385.2	\$394.7	\$384.9	\$318.5
Agglomeration-related variables							
Localisation ($\ln[LQ_{it}]$)	0.302	0.721	0.721	0.590	0.686	0.567	0.247
Urbanisation ($-LMA_MS_i$)	-0.002	0.006	0.006	0.005	0.001	0.001	0.001
Geographic concentration of industry [MS_i]	0.026	0.075	0.075	0.020	0.071	0.020	0.026
LMA size ($\ln[E_i]$)	10.98	1.26	1.25	1.13	0.21	0.15	0.15
Industry size ($\ln[E_i]$)	8.78	1.14	1.14	0.17	1.12	0.16	0.32
LMA market thin-ness (LMA_herf_i)	0.004	0.007	0.007	0.006	0.002	0.002	0.002
Industry non-competitiveness ($Herf_i$)	0.022	0.061	0.061	0.013	0.060	0.013	0.019
Firm variables							
Export [1=yes]	0.288	0.453	0.453	0.367	0.446	0.363	0.226
Birth [1=yes]	0.036	0.187	0.187	0.184	0.187	0.184	0.174
Death [1=yes]	0.041	0.199	0.199	0.197	0.199	0.197	0.189
Seasonal firm [1=yes]	0.035	0.185	0.184	0.182	0.185	0.182	0.175
Firm spans industries [1=yes]	0.009	0.096	0.096	0.093	0.096	0.093	0.097
FTE employment	598	1349	1348	707	1322	699	263
Firm age (topcoded at 6 years)	5.23	1.41	1.40	1.32	1.40	1.31	0.78

Notes: The table is based on a 65 percent sample of data. All statistics are weighted by FTE employment. Columns (a) to (e) report within-group standard deviations, calculated as the mean square error from regressions of the row-variable on dummy variables that define the groups (eg: the underlying regressions for column (a) contain period dummies).

Column (a) of Table 4 reports within-period standard deviations.³⁰ They are very similar to the reported standard deviations from the pooled sample, suggesting that time variation is not a significant driver of overall variation. Controlling for heterogeneity across industries in column (b) substantially reduces the variability of the measures that are observed at an industry level (MS_i , $\ln E_i$, and $Herf_i$). The within industry variation in these variables reflects solely changes over time. Time variation appears to be only about a quarter of the overall variation, suggesting that there is substantial heterogeneity across industries. Although much reduced, the variation is still moderately strong. Variability of other variables is less affected by industry controls.

Within-LMA variation is summarised in column (c).³¹ As expected, the greatest impact of removing between-LMA variation is on measures that are observed only at the area level (LMA_MS_j , LMA_Herf_j , $\ln E_j$). For these variables, within LMA variation over time is only about a fifth as strong as overall variability. The variation in localisation, as measured by the log of the locational quotient, appears to be more affected by removing industry variation than by removing LMA variation, suggesting that there are greater differences across industries than across locations.

Controlling for period, industry and LMA main effects (column (d)) yields estimates that are similar to the within industry variation for industry-level variables and similar to within LMA variation for LMA-level variables. The impact on variability of other variables is modest. Not surprisingly, variability of firm-level variables is more significantly reduced when we control for firm, LMA, and period effects. In this case, the remaining variation is across time, which includes variation due to firms changing industry or changing their distribution across LMAs. Column (d) most closely reflects the variation that is used to identify estimates in the full regression model specifications that follow, although Table 5 and Table 6 report estimates with less stringent controls.

4.2 Agglomeration and productivity - Regression analysis

Table 5 reports regression results that control for period effects, but not for industry, LMA, or firm fixed effects. The first column contains estimates from seven separate regressions, each of which is a regression of average labour productivity on a single covariate and a set of period dummies. These results summarise bivariate relationships, adjusted for aggregate changes over time. All coefficients are statistically significant at 1%.

Firms in own-industry agglomerations are relatively more productive ($\ln LQ_{ij}$), consistent with the effects of localisation, as are industries that are more localised (MS_i). Urbanisation ($-LMA_MS_j$), however, is associated with *lower*

³⁰ For each variable, these are calculated as the standard deviation of the error term from a regression of the variable on a full set of period dummies. Subsequent within-group standard deviations are calculated similarly, but with different sets of dummy variables.

³¹ The underlying regression contains not LMA dummies, but shares of employment in each LMA – location is observed for geographic units rather than enterprises and enterprises can operate in more than one location. LMA employment shares sum to one.

productivity, contrary to the predictions of Jacobs, and the bivariate relationship shown in Figure 3(c). Larger LMAs are more productive ($\ln FTE_j$), although industry size is a disadvantage ($\ln FTE_i$). Industry non-competitiveness ($Herf_i$) is associated with higher productivity, whereas local monopsony power (LMA_Herf_j) is associated with lower productivity. All of these effects are as expected, with the exception of urbanisation effects, which are unexpectedly negative. Measures with an i subscript relate to 4-digit ANZSIC industries. A j subscript denotes a Labour Market Area.

These measures are correlated with each other, and column (2) shows coefficient estimates from a regression where they are all entered simultaneously. The magnitudes change, but the signs of the coefficients remain the same, and all except industry size remain significant. Notably, geographically concentrated industries (MS_i) are more productive, even controlling for the direct effect of localisation ($\ln LQ_{ij}$).

Columns (3) and (4) show the relationship between productivity and the firm-level characteristics that we include. The only difference between the two columns is the inclusion of an exporting dummy in column (4), which has a significant effect on the estimated impact of firm size. As suggested by Figure 4, there is a positive relationship between firm size and productivity. Firms employing over 100 FTE workers have productivity that is, on average, \$35,600 higher than firms of 0 to 5 FTE workers. The inclusion of the export indicator reduces the estimated impact of firm size. Controlling for export activity, productivity is similar, and possibly declining slightly with size, for firms employing between 5 and 100 workers. The productivity premium for the largest firms is roughly halved, to \$16,100. The higher productivity of larger firms appears to be largely due to the fact that larger firms are more likely to export, and exporting is associated with productivity that is substantially higher, by \$52,600. There is a relatively stable positive relationship between firm age and productivity. The 'birth' dummy variable captures firms that are in their first year of existence, and is negative, reflecting the understatement of productivity for these firms, as well as possibly lower productivity of start-ups.³² The omitted firm age category is firms that are 1 to 2 years old. More mature firms have higher productivity than the omitted category although the gradient is slight once firms have reached two years.

Column (5) reports estimates from a regression that includes both concentration and firm-level variables. The estimates are qualitatively similar to those from regressions where these are entered separately, suggesting that the correlation between the two sets of effects is not high.

³² Firms in their final year of existence are captured by the 'death' dummy. The negative coefficient reflects the understated productivity measure.

Table 5 Geographic concentration indices and firm characteristics separately

Lab. Prod.(<i>ijk</i>)	(1)	(2)	(3)	(4)	(5)
Localisation (ln)LQ(<i>ijt</i>)	17.1 (2.5)**	14.1 (2.3)**			7.4 (2.2)**
Urbanisation -LMA MS(<i>jt</i>)	-569.6 (211.9)**	-2,164.3 (313.8)**			-2,208.0 (311.0)**
Geo Conc of Industry MS(<i>it</i>)	292.6 (19.2)**	199.8 (25.6)**			171.5 (27.2)**
LMA size lnLMA FTE(<i>jt</i>)	11.1 (0.7)**	11.0 (1.1)**			8.5 (1.0)**
Industry Size (ln)Ind. FTE(<i>it</i>)	-11.1 (1.5)**	-2.3 (1.7)			-2.4 (1.5)
LMA Thinness LMA Herf(<i>jt</i>)	-1,536.2 (118.6)**	-860.5 (143.1)**			-1,052.0 (151.3)**
Industry Non-compet Herfindahl(<i>it</i>)	250.9 (46.5)**	154.8 (44.1)**			102.3 (41.6)*
Size: >5-10 FTEs			5.2 (1.0)**	3.0 (1.0)**	4.3 (1.0)**
Size: >10-20 FTEs			9.3 (1.3)**	4.9 (1.4)**	5.0 (1.5)**
Size: >20-50 FTEs			11.4 (1.4)**	2.8 (1.5)	2.1 (1.7)
Size: >50-100 FTEs			18.0 (2.4)**	2.5 (2.6)	1.0 (2.8)
Size: >100 FTEs			35.6 (3.5)**	16.1 (3.7)**	15.0 (3.4)**
Age: 2 years			19.2 (3.4)**	19.2 (3.1)**	19.9 (3.4)**
Age: 3 years			18.8 (3.6)**	18.6 (3.2)**	19.3 (3.6)**
Age: 4 years			23.5 (3.7)**	23.2 (3.4)**	24.0 (3.7)**
Age: 5 years			23.0 (4.6)**	22.6 (4.3)**	24.9 (4.6)**
Age: 6+ years			30.8 (3.6)**	27.7 (3.2)**	29.5 (3.6)**
Exporting				52.6 (4.0)**	39.0 (3.8)**
Birth dummy			-11.3 (1.8)**	-12.0 (1.8)**	-14.0 (1.8)**
Death dummy			-20.6 (2.2)**	-21.6 (2.4)**	-22.4 (2.2)**
Seasonal dummy			-19.3 (3.7)**	-16.8 (4.1)**	-15.7 (3.9)**
Multi-industry firm			-6.1 (8.3)	-11.8 (9.2)	-13.3 (8.4)
Period FE	X	X	X	X	X
Observations	1,450,528	1,450,528	1,450,528	1,450,528	1,450,528
R-squared	all 0.00	0.01	0.01	0.01	0.01

Notes: Robust standard errors in parentheses, * significant at 5%; ** significant at 1%
 Dependent variable is “under-represented labour productivity” measured in \$000s.
 Column 1 contains estimates from 7 separate regressions. All regressions are weighted by FTE employment.

Table 6 Controlling for industry, LMA, and enterprise effects

Lab. Prod.(<i>ijkt</i>)	(5) Yr	(6) Yr & Ind	(7) Yr & LMA	(8) Yr, Ind & LMA	(9) Yr, LMA & Firm
Localisation (ln)LQ(<i>ijt</i>)	7.4 (2.2)**	7.8 (2.0)**	9.1 (2.3)**	8.2 (2.1)**	8.7 (4.6)
Urbanisation -LMA MS(<i>jt</i>)	-2,208.0 (311.0)**	-1,047.4 (147.9)**	612.3 (500.1)	606.8 (398.3)	820.5 (354.4)*
Geo Conc of Industry MS(<i>it</i>)	171.5 (27.2)**	144.5 (91.9)	152.0 (27.3)**	141.9 (91.6)	45.1 (57.3)
LMA size lnLMA FTE(<i>jt</i>)	8.5 (1.0)**	7.5 (0.6)**	17.5 (4.3)**	11.5 (3.7)**	4.6 (3.5)
Industry Size (ln)Ind. FTE(<i>it</i>)	-2.4 (1.5)	-14.6 (7.1)*	-3.2 (1.5)*	-14.4 (7.1)*	-7.9 (2.9)**
LMA Non-compet LMA Herf(<i>jt</i>)	-1,052.0 (151.3)**	-195.6 (77.9)*	-409.2 (287.3)	10.7 (251.7)	-77.9 (193.3)
Industry Non-compet Herfindahl(<i>it</i>)	102.3 (41.6)*	157.7 (163.4)	114.2 (43.1)**	159.8 (163.3)	-0.3 (96.8)
>5-10 FTEs	4.3 (1.0)**	4.7 (1.1)**	4.3 (1.0)**	4.7 (1.1)**	-9.5 (2.3)**
>10-20 FTEs	5.0 (1.5)**	6.4 (1.6)**	4.9 (1.5)**	6.4 (1.6)**	-22.8 (3.3)**
>20-50 FTEs	2.1 (1.7)	2.4 (1.8)	2.2 (1.7)	2.5 (1.8)	-35.0 (4.8)**
>50-100 FTEs	1.0 (2.8)	-1.8 (2.9)	-0.3 (2.9)	-2.4 (3.0)	-56.8 (6.7)**
>100 FTEs	15.0 (3.4)**	12.8 (2.9)**	11.3 (3.2)**	12.2 (3.0)**	-79.7 (8.8)**
2 years	19.9 (3.4)**	15.9 (4.3)**	20.7 (3.0)**	16.3 (4.0)**	21.3 (2.7)**
3 years	19.3 (3.6)**	14.2 (4.5)**	20.3 (3.2)**	14.7 (4.2)**	20.8 (3.5)**
4 years	24.0 (3.7)**	18.2 (4.6)**	24.9 (3.4)**	18.7 (4.3)**	23.7 (3.8)**
5 years	24.9 (4.6)**	19.4 (5.3)**	26.1 (4.3)**	20.1 (5.0)**	20.5 (4.2)**
6+ years	29.5 (3.6)**	26.2 (4.5)**	30.6 (3.2)**	26.7 (4.2)**	20.0 (4.7)**
Exporting	39.0 (3.8)**	47.9 (3.7)**	39.8 (3.9)**	47.7 (3.6)**	25.0 (3.8)**
Birth, death, seasonal, multi-industry	X	X	X	X	X
Period FE	X	X	X	X	X
Industry FE		X		X	
LMA FE			X ^(a)	X ^(a)	X ^(a)
Enterprise FE					X
Observations	1,450,528	1,450,528	1,450,528	1,450,528	1,450,528
R-squared	0.01	0.05	0.01	0.05	0.37

Notes: Robust standard errors in parentheses, * significant at 5%; ** significant at 1%

Dependent variable is “under-represented labour productivity” measured in \$000s.

All regressions are weighted by FTE employment.

(a) LMA fixed effects are included as a set of share variables measuring the share of a firm’s FTE employment in each LMA.

The results in Table 5 are derived from pooled cross-sectional estimates, and will thus reflect the effects of heterogeneity across locations, industries and firms. Estimated agglomeration effects may be a result of industry, area, or firm productivity factors that are correlated with, but not a result of, agglomeration. In order to more credibly identify the causal effect of

agglomeration, we wish to isolate the relationships based on variation within industries, within LMAs, and for firms across time.

To this end, Table 6 examines the strength of the link between productivity and agglomeration within industries, within LMAs, and for firms across time. The first column (column (5)) repeats the estimates from the final column of Table 5, which include concentration variables, firm characteristics, and period dummies. Column (6) reports estimates from a regression that also includes dummy variables for each of 407 4-digit industries. The estimated effects thus exclude the impact of industry differences, and reflect the relationship between agglomeration and productivity for firms within an industry, net of overall period effects. We are thus removing, from the estimates in column (5), the (potentially causal) long-term relationship between an industry's agglomeration patterns and its productivity. It may be that some industries are more productive because they are more geographically concentrated. However, this cross-sectional relationship may also reflect other industry-related factors that are correlated with both concentration and productivity, and cannot therefore be interpreted as causal. The industry fixed effects estimates address the more demanding question of whether within-industry variation over time and location in an industry's agglomeration patterns is associated with variation in productivity.

As would be expected as a result of the reduced variation reported in Table 4, the precision of estimates for industry-level variables is lowered when we include industry fixed effects. Changes over time in the degree of industry concentration (MS_i) and industry non-competitiveness ($Herf_i$) are still positively related to industry productivity, but the effects are not statistically significant. There is a statistically significant negative relationship between industry growth ($\ln FTE_i$) and industry productivity, possibly reflecting increased competition, whereby net growth in the number of competing firms may lower average margins, at least in the short term. Own-industry agglomeration ($\ln LQ_{ij}$) continues to be positively and significantly related to productivity, although this could be capturing LMA-wide productivity effects in LMAs where own-industry agglomerations occur.

Estimates that control for mean differences across LMAs are reported in column (7). These estimates exclude any stable differences in productivity across LMAs, some of which may be related to agglomeration characteristics of the LMA. The presence of productive local amenities such as transport infrastructure or favourable climate may produce a positive relationship between agglomeration and productivity, and are an important factor in the analysis of agglomeration effects. The LMA fixed effects estimates, however, exclude such factors, and for LMA-level measures, identify the link between productivity and changes over time in LMA agglomeration characteristics. As with the impact of industry fixed effects on industry variables, the LMA fixed effects reduce the precision of estimates of the effect of LMA variables. The most notable impact of controlling for stable LMA differences is that the negative relationship between LMA diversity ($-LMA_MS_i$) and productivity is reversed, and loses significance. More diverse LMAs are thus *less* productive in the cross section, but increases in urbanisation over time are associated with higher productivity growth. It is likely

that the cross-sectional relationship is reflecting other LMA factors, and not a causal effect of urbanisation *per se*. LMA size ($\ln FTE_j$) remains positively related to productivity, suggesting generalised agglomeration scale effects that are not due solely to diversity or own-industry concentration.

Including both industry and LMA fixed effects, as is done in column (8) corrects the industry fixed effects estimates for the influence of LMA heterogeneity and corrects the LMA fixed effects estimates for the influence of industry heterogeneity. It would appear that these inter-relationships are not great, given the similarity of the column (8) coefficients to the estimates of within-industry effects of industry variables in column (6) and of within-LMA effects of LMA measures in column (7).

The final column of Table 6 strengthens the test for a causal relationship between agglomeration and productivity by focusing solely on variation over time for a firm. Changes over time in the concentration measures can reasonably be thought of as exogenous for a given firm.³³ The estimates in column (9) also remove the influence of any sorting of more productive firms into concentrated industries, diverse LMAs, or into areas of own-industry agglomeration.³⁴ The point estimate of the positive impact of localisation does not change markedly, although the estimate just loses significance ($p=0.059$). In contrast, the estimated positive effect of urbanisation strengthens, and becomes significant at a 5% level.

The estimated industry-wide effect of localisation, as captured by MS_{it} , is reduced, and is insignificant. The productivity advantages of LMA size are reduced in magnitude and significance, but the effect of industry size remains negative and significant, which we tentatively interpret as the effect of increased competition. Finally, the productivity effects of local monopsony (LMA_Herf_{jt}) and industry monopoly ($Herf_{it}$) become smaller and remain insignificant once unobserved firm heterogeneity is controlled for.

The firm fixed-effects have a pronounced effect on the estimated relationship between productivity and firm-level variables. The estimated impact of exporting reflects productivity changes for firms entering or exiting the export market. The effect is reduced, to \$25,000, but remains highly significant. The firm-size productivity gradient becomes strongly negative. We believe that this reflects the impact of measurement error in our annual employment measure. In periods when a firm's employment is overstated, productivity is understated, due to the use of employment as the denominator in the productivity measure.³⁵

³³ For multi-unit enterprises, variation in agglomeration measures can also arise because of (possibly endogenous) changes in the distribution of FTE activity across LMAs. This is a result of the within-enterprise averaging of agglomeration measures, as described in the data section.

³⁴ The within-firm variation in concentration measures includes variation resulting from firms changing industry or changing the distribution of their employment across LMAs over time.

³⁵ The correspondence is not exact. The employment measure used in the calculation of labour productivity is an average of initial and final levels, whereas the categorical size covariates are based on initial size.

Given the relatively small variation in FTE employment within firms,³⁶ measurement error that is small relative to total variation may account for a large proportion of within-firm variation.

Although some of the estimated effects lose statistical significance in the presence of fixed effects, it may be that within-firm variation over 10 years is insufficiently large to allow us to identify the effect with sufficient precision. For instance, the coefficient on own-industry agglomeration ($\ln LQ_{ij}$) remains relatively stable across specifications, even though the increased standard errors render it insignificant.

Overall, the firm fixed-effects and industry-LMA fixed effects results both provide some support for the operation of both Marshallian localisation effects, through own-industry agglomeration and industry concentration, and for Jacobs urbanisation effects, operating through the diversity of industry mix within a local labour market.

There remains the question of whether the estimated effects are economically significant. To gauge this, we report in Table 7 the estimated impact of the observed changes in each of the concentration measures. The first column of Table 7 contains the estimated coefficients from column (8) of Table 6. Multiplying these coefficients by the observed decadal change in each agglomeration measure, as shown in the second column, gives an estimate of the contribution of agglomeration change during the study period. The implied partial effects on labour productivity are shown in the third column. Over the 10-year study period, average labour productivity grew by 33 percent. The final two columns express the partial effects as a proportion of the mean productivity level, and of the actual change in productivity respectively.

Table 7: Economic significance of estimated effects

	Coefficient	Actual change (%)	Partial effect (b*dx)	% of mean productivity	% of productivity change
Localisation[$\ln LQ_{ij}$]	8.2	-0.030 (-9%)	-\$300	-0.4%	-1.3%
Urbanisation[-LMA_MSj]	606.8*	-0.001 (-54%)	-\$600	-1.1%	-3.3%
Geographic Concentration[MSi]	141.9	0.004 (17%)	\$600	1.0%	3.1%
LMA Size[$\ln E_j$]	11.5	0.314 (3%)	\$3,600	6.4%	19.5%
Industry Size[$\ln E_i$]	-14.4**	0.250 (3%)	-\$3,600	-6.4%	-19.5%
LMA market thinness (LMA_{herfj})	10.7	-0.001 (-18%)	-\$0	-0.0%	-0.1%
Industry non-compet[Herf]	159.8	-0.007 (-26%)	-\$1,100	-2.0%	-6.1%

Notes: Asterisks indicate the significance of the regression coefficient: * significant at 5%; ** significant at 1%; The partial effect is calculated using the regression coefficients from column (8) of Table 6.

³⁶ Table 4 shows an overall standard deviation of firm size of 1,349 and a within-firm variation of only 263. This implies that the within-firm variance is only around 4 % of total variance.

Industry size and LMA size make the largest absolute contributions, coincidentally of the same magnitude, though opposite signs, of -\$3,600 and \$3,600 respectively. The contribution of industry size is negative because of a negative regression coefficient. Similarly, changes in urbanisation make a moderate-sized negative contribution of \$600 because the reduction in diversity that occurred reduced predicted productivity. The degree of localisation ($\ln LQ_{ij}$) declined over the period, leading to a small negative contribution of -\$300. Changes in industry-level localisation (MS_i) yielded a small increase of \$600 in mean productivity.³⁷

Overall, although the effects of location measures are not all estimated with statistical precision, the magnitude of estimated effects is large enough to be of economic interest, although some are modest. Changes in industry and LMA size have the largest absolute effects on productivity, of -6 percent and 6 percent respectively. These changes account for -19 and 19 percent of observed productivity change respectively. The reduction in urbanisation leads to a lowering of productivity growth by 1 percent and the reduction in localisation leads to a slight (0.4%) lowering of productivity growth.

4.3 Variation in agglomeration effects

The overall impact of agglomeration variation is estimated to be modest, yet it is possible that agglomeration forces operate more strongly for particular industries or types of firm. Maré (2005) found that only about 30 percent of employment is in geographically concentrated industries, suggesting that localisation and urbanisation effects may have restricted industrial scope.

Previous studies suggest that urbanisation may be particularly important for small and new firms, whereas more mature firms benefit more from greater localisation (Holmes and Stevens (2002), Duranton and Puga (2001)). Henderson (2003) also shows that localisation is particularly valuable for manufacturing firms without corporate affiliates. Nakamura (1985) and Henderson (1986) estimate that localisation effects are more important than urbanisation for manufacturing production generally. Henderson suggests, however, that urbanisation may be more relevant for firms in more ‘experimental’ industries – a conjecture supported by Ó hUallacháin (1989), who finds that urbanisation rather than localisation is a key factor for at least some service industries.

Table 8 contains a range of estimates similar to the ‘base-case’ regression shown in column 8 of Table 6, but for a range of subsets of the data. The aim is to highlight the variation in and robustness of the estimated productivity effects of agglomeration across dimensions that have been highlighted in the literature.

The first row of Table 8 reproduces column (9) from the previous table, as a point of reference for subsequent estimates. The first split of the data is

³⁷ . For comparison, the impact of observed changes in export activity (a decadal increase of 2.2 percentage points) are estimated to have raised average labour productivity by around \$600, or about 1 percent over the decade.

between public and private sectors, shown in the second and third rows. This highlights a sharp distinction in the size of estimated effects. The estimated agglomeration effects are all more modest for the private sector than for the public sector. The estimated effect of localisation is large and significant for public sector firms, but we are cautious in interpreting this result given the difficulties in measuring value added in the public sector, and the peculiarities of a sector that is highly concentrated in the capital city, Wellington.³⁸ The remainder of Table 8 presents analyses of subsets of private sector firms only, which account for just under 90 percent of observations.

For private sector firms overall, the only dimension of agglomeration for which there is a statistically significant positive relationship with firm productivity is the size of the local labour market ($\ln FTE_{jt}$). The effect of neither localisation nor urbanisation is significant, although both have positive point estimates.

For smaller firms, the benefits of localisation appear to be greater, as shown by the significant localisation coefficient (3.7) from an unweighted regression, shown in the fourth row of the table, in which large-FTE firms receive less weight. The firm size pattern is confirmed by examining separate estimates by firm size category. As shown in Table 8, localisation is negatively, though insignificantly related to productivity for the largest firms (FTE>100). For smaller sized firms, the benefits of agglomeration appear to rise with firm size. The smallest positive benefit (2.8) accrues to firms with 0 to 5 FTE employment.

For firms between 5 and 50 FTE, the coefficient is 5.6, rising to 7.7, but losing significance, for firms between 50 and 100 FTE. This pattern is contrary to the patterns observed in previous studies, whereby localisation is particularly important for small and young firms. These previous findings were, however, predominantly for manufacturing firms, whereas our estimates reflect effect across a more representative population of industries.

Our findings do, however, support previous findings that the productivity benefits of urbanisation are strongest for young firms. In particular, for firms that have operated for only 1 to 2 years, urbanisation has a strong, statistically significant positive impact on labour productivity. In contrast, localisation has the strongest positive effect on firms that have operated for 5 or more years. This is consistent with the life-cycle pattern described in Duranton and Puga (2001). One possible reason for this life-cycle pattern not showing up in the firm size profile is that New Zealand contains many well-established small firms, so that the results by firm size represent a mix of age and size effects.

³⁸ . The public sector regression is the only one for which exporting is not associated with higher productivity, even though the proportion of employment in public sector exporting firms (31%) is similar to the proportion for the sample as a whole.

Table 8 Variation in the nature and strength of agglomeration effects – size, age, and exporting

	Localisation (ln)LQ(<i>ijt</i>)		Urbanisation -LMA MS(<i>jt</i>)		Geo Conc of Ind MS(<i>it</i>)		LMA size lnLMA FTE(<i>jt</i>)		Industry Size (ln)Ind. FTE(<i>it</i>)		LMA Non-compet LMA Herf(<i>jt</i>)		Ind Non-compet Herfindahl(<i>it</i>)		N (000)
Base Case (8)	8.2	(2.1)**	606.8	(398.3)	141.9	(91.6)	11.5	(3.7)**	-14.4	(7.1)*	10.7	(251.7)	159.8	(163.3)	1,450
Public Sector	47.6	(13.9**)	1,292.7	(1,552.8)	1,246.0	(655.7)	14.3	(12.1)	11.0	(17.6)	511.1	(504.6)	1,452.3	(780.2)	58
Private Sector	1.2	(1.7)	238.1	(292.5)	67.0	(85.9)	7.3	(2.2)**	-10.6	(8.7)	-440.9	(236.0)	30.6	(100.8)	1,279
Private Sector only unweighted	3.7	(0.8)**	243.6	(150.6)	23.1	(85.4)	4.5	(4.6)	7.2	(7.0)	-18.7	(83.6)	188.7	(188.5)	1,279
Firm Size															
(0,5]	2.8	(0.6)**	140.9	(87.2)	-51.3	(51.4)	0.9	(4.4)	-0.9	(7.5)	52.9	(50.7)	291.6	(263.8)	1,111
(5,10]	5.6	(1.8)**	-239.3	(309.4)	33.2	(216.4)	-4.0	(3.5)	15.0	(17.8)	10.4	(96.5)	12.3	(271.0)	94
(10,20]	5.6	(2.2)*	-216.5	(304.8)	132.4	(166.5)	-6.1	(4.1)	-11.7	(9.2)	94.1	(149.8)	-315.3	(162.0)	43
(20,50]	5.6	(2.8)*	-354.4	(470.8)	-71.1	(114.2)	-3.1	(3.7)	-1.3	(9.0)	-418.6	(407.5)	-297.7	(326.4)	22
(50,100]	7.7	(4.6)	-1,087.8	(1,618.1)	39.3	(103.0)	6.9	(6.8)	-26.7	(20.1)	-286.5	(1,375.5)	242.3	(393.9)	5
100+	-9.7	(7.8)	1,808.9	(1,591.4)	-117.5	(113.9)	3.8	(5.7)	-27.9	(35.3)	-777.8	(1,051.8)	-170.7	(100.5)	5
Age of firm															
up to 1 year	1.6	(1.9)	199.3	(423.5)	26.8	(134.7)	-14.2	(14.3)	-3.6	(9.9)	-536.5	(254.1)*	-808.2	(493.2)	125
1-2 years	2.4	(2.2)	1,695.5	(840.7)*	-260.0	(175.0)	20.3	(18.3)	-27.0	(22.3)	381.7	(482.9)	691.5	(942.7)	173
2-3 years	0.7	(1.6)	319.9	(311.4)	115.6	(133.0)	-16.4	(11.2)	-18.0	(9.5)	290.3	(226.1)	-161.6	(169.9)	158
3-4 years	2.2	(1.9)	-483.4	(394.3)	228.3	(219.4)	-3.0	(9.0)	-1.9	(12.4)	64.3	(228.8)	115.0	(247.3)	131
4-5 years	2.8	(2.1)	-369.8	(670.2)	324.7	(201.6)	-0.8	(5.3)	15.3	(10.2)	-962.8	(746.2)	46.2	(411.2)	107
5 or more years	4.2	(1.7)*	447.6	(363.8)	57.4	(105.1)	5.8	(2.2)**	-12.0	(14.1)	-356.7	(267.3)	65.8	(148.3)	585
Export Activity															
Never exports	2.1	(0.4)**	229.3	(117.6)	75.5	(37.8)*	7.8	(6.2)	-3.8	(4.8)	11.4	(79.1)	-88.8	(51.0)	1,014
Sometimes exports	3.8	(2.9)	525.8	(846.3)	-156.6	(88.2)	-13.3	(6.1)*	-0.0	(24.0)	898.2	(505.7)	316.3	(185.9)	219
Always exports	11.7	(4.7)*	-537.4	(1,468.5)	17.7	(153.4)	5.3	(3.3)	-10.6	(19.0)	-2,521.0	(1,064.3)*	-71.0	(122.6)	46
Multi-LMA operation															
single-LMA	3.9	(0.7)**	48.8	(250.3)	22.8	(63.3)	8.4	(10.2)	-7.3	(6.6)	-242.3	(231.8)	-55.2	(126.9)	896
multi-LMA	2.2	(2.9)	563.8	(560.0)	44.2	(121.8)	5.6	(2.4)*	-12.2	(14.6)	-525.3	(377.5)	170.2	(122.8)	383

Notes: Robust standard errors in parentheses, * significant at 5%; ** significant at 1%

Dependent variable is “under-represented labour productivity” measured in \$000s.

All regressions are weighted by FTE employment.

(a) LMA fixed effects are included as a set of share variables measuring the share of a firm’s FTE employment in each LMA.

Exporting firms are less reliant on the size of the local output market than are non-exporting firms, although advantages of thick input markets affect both types of firm. Exporting firms tend to produce at greater scale, and we might therefore expect the benefits of localisation to be strongest for exporting firms. The ‘export activity’ panel of Table 8, however, shows no statistical difference in agglomeration effects between firms that always export, firms that never export, and firms that switch in or out of exporting.

The final panel of Table 8 compares results for firms that operate in a single LMA with results for firms that operate across LMAs. This is in part a robustness check, given the averaging of agglomeration measures within multiple-LMA firms, as described above. For single-LMA firms, changes in agglomeration measures are more clearly exogenous compared with those for multiple-LMA firms, which are able to reallocate activity across locations. There are no statistically significant differences in estimated effects across these two groups.

Identifying industry variation in agglomeration effects has been a key feature of previous studies of the relationship between agglomeration and productivity, although many studies have been restricted to manufacturing industries only. Our data allow us to examine effects across a broad range of industries. Results are summarised in Table 9, first by primary, secondary and service sectors, then by 1-digit industry groupings, and then by derived groupings of 4-digit industries based on similarity of locational distribution.

Dividing our sample up broadly into primary, secondary, and services firms,³⁹ we find that firms in secondary industries, which includes manufacturing, display the least significant effect of localisation. Although urbanisation advantages appear to be large for secondary industries, the impact is statistically insignificant. LMA size is the only significant positive agglomeration influence on secondary firms. For primary firms, localisation is a disadvantage (coefficient of -7.7), as might be expected for industries that are more dispersed, as a consequence of their necessary proximity to natural resources. Primary industries dominated by relatively few firms (high $Herf_{it}$) also appear to have relatively high labour productivity.

The strongest positive effect of localisation is found for firms in service industries (5.5). Local diversity is often argued to be particularly important for creative and knowledge’ industries, which are concentrated in services. Perhaps surprisingly, the effect of urbanisation is estimated to be slightly negative though insignificant. The effect of local market size is positive, although not statistically significant for service industries.

³⁹. The ‘primary’ sector includes agricultural industries, and mining and quarrying (ANZSIC divisions A and B). The secondary sector includes ANZSIC divisions C, D and E (Manufacturing, Electricity, Gas & Water, and Construction). The remaining industries are classified as tertiary.

Table 9 Variation in the nature and strength of agglomeration effects – Industry variation

	LnLQ		LMA_MS		MS		FTEj		FTEi		LMA_Herf		Herf		N (000)
Sector															
Primary	-7.7	(3.5)*	644.8	(419.2)	135.6	(257.0)	23.3	(16.5)	-28.3	(30.5)	179.7	(225.2)	1,100.0	(403.1)**	51
Secondary	-0.4	(2.6)	753.6	(733.8)	206.9	(156.2)	8.0	(2.8)**	-4.8	(14.1)	-713.7	(538.0)	73.7	(107.9)	329
Services	5.5	(2.2)*	-91.9	(235.0)	-69.7	(86.9)	6.4	(3.5)	-15.3	(10.6)	-163.8	(134.6)	-146.3	(231.9)	899
Industry															
A: Agriculture	-8.5	(3.4)*	241.8	(464.8)	-72.3	(243.5)	2.3	(52.0)	24.7	(17.2)	447.9	(200.3)*	2,399.7	(1,142.2)*	49
C: Manufacturing	0.8	(2.3)	365.7	(991.1)	229.5	(146.5)	7.4	(2.5)**	-7.9	(11.0)	-1,163.3	(699.6)	19.1	(95.4)	119
E: Construction	-1.2	(1.2)	112.3	(123.5)	-152.7	(90.1)	8.4	(5.6)	5.3	(6.0)	69.5	(54.3)	-426.7	(423.1)	209
F: Wholesale	41.2	(12.3)**	9,411.5	(4,791.3)*	183.4	(438.2)	39.8	(32.9)	-155.1	(41.9)**	6,178.8	(2,075.8)**	4,806.5	(2,296.2)*	97
G: Retail	-0.7	(0.7)	65.9	(76.1)	4.6	(126.0)	2.5	(3.8)	6.7	(9.3)	8.6	(47.2)	840.1	(642.3)	193
H: Accommod.	-2.2	(0.9)*	289.5	(189.4)	297.4	(177.3)	0.2	(6.1)	-17.5	(9.4)	-147.1	(107.6)	-222.1	(1,066.0)	51
I: Transport	8.9	(5.5)	329.6	(800.1)	260.4	(212.4)	45.3	(20.0)*	-20.4	(26.5)	-521.1	(476.7)	-428.5	(267.1)	64
J: Communication	-24.5	(16.7)	-8,217.2	(6,448.8)	-1,378.2	(1,500.4)	-64.7	(342.5)	-32.5	(149.6)	-1,741.1	(2,065.9)	-523.4	(4,162.8)	20
K: Financial & Insurance	45.1	(52.5)	-13,849.4	(8,877.5)	-61.1	(268.5)	-15.0	(22.2)	96.5	(106.9)	-3,461.6	(2,884.9)	548.3	(496.7)	20
L: Property & Business Services	7.4	(7.8)	-1,004.8	(445.7)*	73.2	(94.9)	3.2	(2.2)	-3.4	(5.3)	-381.8	(470.6)	-35.1	(153.6)	311
O: Health Services	-1.2	(1.3)	109.2	(495.8)	31.7	(30.9)	-4.2	(2.5)	-8.6	(4.2)*	116.3	(217.9)	-30.5	(29.0)	55
P: Cultural & Personal Services	-2.4	(4.1)	772.7	(677.4)	-99.6	(101.8)	-4.6	(7.6)	-12.8	(4.7)**	207.8	(359.7)	34.9	(180.5)	34
Q: Personal Services	-0.4	(2.1)	48.3	(185.6)	185.1	(77.7)*	-4.3	(9.6)	7.1	(7.2)	-54.4	(80.1)	-13.0	(78.2)	45
Co-location Groups															
Concentration Manufacturing	7.9	(3.4)*	5,490.4	(2,332.3)*	-116.8	(78.7)	-0.6	(4.6)	-67.1	(15.9)**	-289.8	(1,763.3)	175.5	(214.3)	60
Wholesale	1.8	(8.8)	-4,047.7	(2,060.4)*	282.5	(127.6)*	18.5	(9.1)*	-28.0	(54.1)	-1,856.1	(2,108.4)	-1,106.1	(280.2)**	82
Business Services	-0.9	(4.2)	-1,555.1	(834.3)	-175.7	(101.5)	0.1	(3.3)	12.6	(15.3)	-453.5	(301.7)	48.9	(279.3)	288
Local Services	2.1	(0.8)**	175.8	(147.2)	-66.4	(67.2)	10.1	(16.0)	-1.8	(3.8)	-11.3	(73.1)	-81.9	(88.6)	690
Local Manufacturing	29.5	(6.9)**	2,319.8	(1,090.5)*	597.7	(459.1)	16.6	(11.9)	-78.7	(28.1)**	1,613.2	(561.9)**	1,905.3	(1,310.8)	256
Resource-based	-8.5	(2.5)**	-70.4	(904.9)	430.3	(132.1)**	12.1	(6.6)	-17.0	(13.9)	-841.6	(637.3)	-321.4	(145.1)*	66

Notes: Robust standard errors in parentheses, * significant at 5%; ** significant at 1%

Dependent variable is “under-represented labour productivity” measured in \$000s.

All regressions are weighted by FTE employment.

(a) LMA fixed effects are included as a set of share variables measuring the share of a firm’s FTE employment in each LMA.

Results by 1-digit industry confirm the patterns for primary and secondary industries, and reveal a diversity of patterns across service industries. Agricultural (division A) industries dominate the Primary sector, so not surprisingly, results are similar for these two groupings.⁴⁰ Results for Manufacturing industries (division C) are likewise similar to the secondary industries grouping, given that they account for over two-thirds of secondary industry employment. There are, however, distinct patterns for the various 1-digit industries within the Services sector. The strong positive effect of localisation on productivity for the Services sector as a whole reflects the relationship for wholesaling industries (division F) and possibly for the property and business services division (L), although the latter is not statistically significant. Wholesaling industries also benefit most from urbanisation, in contrast to Property and Business Services, and possibly Finance and Insurance industries, for which urbanisation is negatively associated with productivity.

One drawback of examining industries at the one-digit level is that this aggregation can conceal within-industry variation in location patterns and the strength of agglomeration effects. This issue is particularly acute for manufacturing industries. Around a quarter of New Zealand's manufacturing industries are linked to primary industries, with a low level of geographical concentration.⁴¹ The remainder of Manufacturing employment is split fairly evenly between a highly dispersed group, and component industries with a high degree of geographic concentration.⁴² Wholesaling is similarly split between highly concentrated industries (55% of division F Employment) and industries that are geographically dispersed (35% of division F employment).

As an alternative to grouping 4-digit industries into 1-digit industry divisions, Table 9 presents results for a grouping of 4-digit industries based on their location patterns. Industries that tend to choose the same locations are grouped together. The grouping of industries is done by first calculating pairwise correlations between location decisions for each pair of 4-digit industries. The derivation of the correlation index is similar to that of the Maurel and Sedillot concentration index. We transform the matrix of pairwise colocation indices into a distance matrix by subtracting each entry from 1, and apply a standard statistical clustering procedure to derive a hierarchical clustering of industries, using Ward's method on unsquared distances. A fuller description of the method and of the resulting groups can be found in Maré (2005).

Table 10 summarises the six 'colocation groups' that are examined in Table 9. The groups are listed in descending order of geographic concentration – the concentrated manufacturing group has the highest value of the Maurel-Sédillot index of concentration and the resource-based group is the most dispersed. The

⁴⁰ . Mining and Quarrying industries (division B) are the only non-Agricultural industries also included in the primary industry grouping, but account for only 2,100 observations, and so are not reported separately.

⁴¹ . The six main industries in this category are 'Meat processing', 'Log sawmilling', 'Seafood processing', 'Fruit and Vegetable processing', 'Pulp, Paper and Paperboard Manufacturing', and 'Plywood and Veneer Manufacturing'.

⁴² . For a fuller discussion, and a listing of concentration patterns for 4-digit industries, see Maré (2005).

names assigned to the groups are indicative only. Each group contains a variety of industries and the titles indicate the most common type of industry in each group.

Table 10: Colocation Groups

Name	# 4-digit industries	mean annual # firms in sample (000)	mean annual FTE in sample (000)
Concentrated Manufacturing	65	6.0	51.9
Wholesale	47	8.2	47.2
Business Services	50	28.8	139.3
Local Services	93	69.0	324.5
Local Manufacturing	81	25.6	155.2
Resource based	41	6.6	61.9

Notes: The list of groups excludes 30 4-digit industries in smaller clusters, collectively accounting for around 1,000 firms and FTE employment of 28,000.

The final panel of Table 9 contains regression estimates for each of these groups separately. The groups appear in the same order as in Table 10 – from most geographically concentrated to most dispersed.

The ‘Concentrated Manufacturing’ colocation group contains the most geographically concentrated industries. (Employment in this group is 50% Manufacturing, 25% Wholesaling, and 12% Transport & Storage.) For this group, both localisation and urbanisation have significant positive effects on productivity. Unexpectedly, both localisation and urbanisation are also significant and positive for the ‘Local Manufacturing’ colocation group, which has its employment geographically dispersed roughly in proportion to total employment. (Employment for this group is 25% manufacturing, with 10-15% in each of Retail, Construction, Wholesaling, and Transport & Storage.) Even though employment is distributed proportionately throughout New Zealand, agglomeration appears to confer benefits even to these industries.

Business Services industries – including mainly components industries ‘Property and Business Services’ and ‘Finance and Insurance’ have a fairly high degree of geographic concentration, but their labour productivity is not significantly linked to agglomeration factors. The point estimates for the impact of localisation and urbanisation are, in fact, both negative. Again, there is little evidence in the data that agglomeration is a major influence on productivity for these knowledge industries.

Resource-based industries, over half of which are manufacturing industries, and a quarter of which are from the Agricultural industry division (a), have lower productivity if they are localised or urbanised. The inclusion of these negative relationships would have reduced the estimated productivity effects of agglomeration for the 1-digit manufacturing industry division. The within-industry variation in agglomeration effects for manufacturing point to the importance of choosing an appropriate industry grouping.

5 Summary and conclusions

The combination of Business Demography (BD) and Business Activity Indicator (BAI) data provides a valuable platform for the analysis of firm productivity and the role played by location, industry, and firm characteristics.

Net sales recorded for New Zealand's GST system provides a credible proxy for firm productivity and the almost universal coverage of combined data makes it particularly attractive as a foundation for research.

A key feature of our empirical approach is the symmetric treatment of location and industry dimensions of agglomeration. This allows us to investigate a range of potential explanations for agglomeration effects. Our approach allows us to simultaneously test for the existence of localisation effects associated with own-industry agglomeration, and for urbanisation effects that operate through the diversity of local industrial mix.

The cross-sectional relationships between labour productivity and agglomeration are strong and statistically significant. Localised firms are more productive, as are firms in large and competitive LMAs. However, local labour market areas (LMAs) that are more diverse have lower average productivity, contrary to what is implied by Jacobs for differences across cities. Local labour market size is a positive correlate of firm productivity, suggesting more general benefits of scale or market thickness.

The patterns are different when we control for stable differences between different industries or LMAs. These differences may reflect genuine agglomeration factors such as the presence of local productive amenities, but the inability to distinguish these from other omitted industry or LMA factors makes cross-sectional variation a weak basis on which to base inferences about the productivity effects of agglomeration. We apply more stringent tests by controlling for firm, location, period, and firm fixed effects. The variation that is used in the firm fixed effects estimates to identify the impact of agglomeration measures is variation over time for a particular firm. Changes over time in patterns of agglomeration are arguably more exogenous to a given firm than is variation across industries or LMAs, which more strongly reflect endogenous location choices. We find some support for a positive productivity effect of changes in both localisation and urbanisation.

The firm fixed-effect estimates are similar to estimates controlling for industry, LMA and time effects only, and the latter are preferred because of the limited within-firm variation, and the possible attenuation bias arising from measurement error. The estimated effects of agglomeration from this preferred specification are not economically significant, with the possible exception of LMA size, which is estimated to have contributed 6% productivity growth over the ten-year study period, with a balancing negative impact of increasing industry size over the same period. Overall, the results provide some support for the operation of both Marshallian agglomeration effects, through own-industry agglomeration (localisation), and for Jacobs-type urbanisation effects, operating through the diversity of industry mix within a local labour market.

We examine variation in the strength of productivity effects of agglomeration for different types of firms and industries. Agglomeration effects are more pronounced for the public sector than for the private sector. Within the private sector, benefits of urbanisation (industry diversity) accrue to young firms, whereas localisation and LMA size boost productivity most for firms that have been in operation for 5 or more years. This pattern is consistent with life-cycle

models of agglomeration effects. In contrast, small firms appear to benefit most from localisation.

Across industries, agglomeration measures are negatively related to productivity for resource-based industries such as agriculture, fishing, and primary processing components of manufacturing. Other manufacturing industries benefit from both localisation and urbanisation, as do wholesaling industries overall. There is no firm evidence of productivity benefits from agglomeration for Service industries other than wholesaling.

One of the motivations for undertaking this study was to contribute to ongoing policy debates on the scope of geographically targeted policies to raise average productivity. Our findings provide tentative support for such policies, but emphasise that the effects are neither economically large in aggregate, nor uniform in their impact across different firms and industries.

References

- Carroll, Nick; Dean Hyslop; David C. Mare; Jason Timmins and Julian Wood. 2002. "An Analysis of New Zealand's Business Demography Database," *New Zealand Economic Papers*, 36:1, pp. 59-61.
- Ciccone, Antonio. 2002. "Agglomeration Effects in Europe," *European Economic Review*, 46:2, pp. 213-27.
- Ciccone, Antonio and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity," *American Economic Review*, 86:1, pp. 54-70.
- Duranton, Gilles and D Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review*, 91:5, pp. 1454-77.
- Duranton, Gilles and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies" in *Handbook of Regional and Urban Economics, Vol 4*, Vernon J. Henderson and J F. Thisse, Eds. Amsterdam: North-Holland.
- Glaeser, Edward L.; Hedi D. Kallal; Jose A. Scheinkman and Andrei Shleifer. 1992. "Growth in Cities," *Journal of Political Economy*, 100, pp. 1126-52.
- Glaeser, Edward L. and David C. Maré. 2001. "Cities and Skills," *Journal of Labor Economics*, 19:2, pp. 316-42.
- Henderson, J. V. 2003. "Marshall's Scale Economies," *Journal of Urban Economics*, 53, pp. 1-28.
- Henderson, Vernon J. 1986. "Efficiency of Resource Usage and City Size," *Journal of Urban Economics*, 19, pp. 47-70.
- Holmes, Thomas J. and John J. Stevens. 2002. "Geographic Concentration and Establishment Scale," *Review of Economics and Statistics*, 84, pp. 682-90.
- Jacobs, Jane. 1969. *The Economy of Cities*, New York: Vintage Books.
- Maré, David C. 2005. "Concentration, Specialisation and Agglomeration of Firms in New Zealand," *Motu Working Paper 05-12*, Motu Economic and Public Policy Research, Wellington, NZ. Available online at http://www.motu.org.nz/motu_wp_series.htm.
- Marshall, Alfred. 1920. *Principles of Economics*, 8th ed., London: Macmillan. Available online at <http://www.econlib.org/library/Marshall/marP1.html>. Last accessed 4 November 2005.
- Maurel, Françoise and Beatrice Sedillot. 1999. "A Measure of the Geographic Concentration in French Manufacturing Industries.," *Regional Science and Urban Economics*, 29:5, pp. 575-604.
- Moomaw, Ronald L. 1998. "Agglomeration Economies: Localization or Urbanization?," *Urban Studies*, 25, pp. 150-61.
- Nakamura, R. 1985. "Agglomeration Economies in Urban Manufacturing Industries: a Case of Japanese Cities," *Journal of Urban Economics*, 17, pp. 108-24.
- Newell, James O. and Kerry L. Papps. 2001. "Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data," *Occasional Paper 2001/6*, Labour Market Policy Group, Department of Labour, Wellington, NZ. Available online at <http://www.dol.govt.nz>.

- Ó hUallacháin, Breandán. 1989. "Agglomeration of Services in American Metropolitan Areas," *Growth and Change*, 20:3, pp. 34-49.
- Rosenthal, Stuart and William Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies" in *Handbook of Regional and Urban Economics, Vol 4*, Vernon J. Henderson and J F. Thisse, Eds. Amsterdam: North-Holland.
- Seyb, Allyson. 2003. "The Longitudinal Business Frame," *Unpublished paper*, Statistics New Zealand, Christchurch. Available online at <http://www.stats.govt.nz/NR/rdonlyres/1C948CE4-BCE4-40E1-B08C-EFE5B05F927D/0/LongitudinalBusinessFrame.pdf>. Last accessed 4 November 2005.
- Statistics New Zealand. 2004. "Information About the Business Demographic Statistics," *Online documentation*, Statistics New Zealand, Wellington. Available online at <http://www.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Demographic+Statistics>. Last accessed 4 November 2005.
- Statistics New Zealand. 2006. "Productivity Statistics (1988–2005) - Hot Off The Press," *Released 28 March 2006*. Available online at www.stats.govt.nz.

Appendix A: Minimum industry coverage

Appendix Table 1: Industry coverage

		Year													Coverage							
		87	88	89	90	91	92	93	94	94	95	96	97	98	99	00	01	02	03	base	9503	9703
A011	Hort&Fruit																					
A012	Grain Sheep Beef																					
A013	Dairy																					
A0141	Poultry Meat																					
A0142	Poultry Eggs																					
A015	Other Livestock																					
A016	Other Crop																					
B1102	Brown Coal Mine																					
B1511	Petrol explor own																					
B1520	Other mining																					
C2240	Clothing Manufacture																					
C2249	???																					
G5110	Supermkt/ groc																					
I6501	Pipeline transp																					
I6701	Grain storage																					
K7324	Money Mkt dealers																					
L7711	Resid. Prop ops																					
M8112	???																					
M8130	Foreign Govt Reps																					
M8200	Defence																					
O8710	Child Care Serv																					
O8722	Resid Care nec																					
O8729	Non resid care nec																					
P9322	Casinos																					
Q9610	Religious Orgs																					
Q9621	Bus&Prof orgs																					
Q9622	Labour Assocs																					
Q9629	Interest gps nec																					
	ALL OTHER																					

Notes: Shaded cells indicate that the ANZSIC industry group is within coverage in the relevant year. The three final columns show which industries are continuously within coverage for the entire period (base), since 1995 (9503) and since 1997 (9703). All analyses in the paper use base coverage.

Appendix Table 2: Impact of Industry Coverage Restrictions

	Percent of geographic units retained	Percent of FTE employment retained
<i>Old BD series</i>		
1987	99.9%	99.9%
1988	99.9%	99.9%
1989	99.9%	99.9%
1990	99.9%	99.9%
1991	99.9%	99.9%
1992	99.9%	100.0%
1993	99.9%	99.9%
1994	99.9%	99.9%
<i>New BD series</i>		
1994	97.9%	95.3%
1995	98.1%	95.5%
1996	97.2%	94.6%
1997	96.0%	94.0%
1998	76.9%	85.6%
1999	96.0%	94.0%
2000	96.1%	94.0%
2001	96.1%	94.0%
2002	96.0%	93.9%
2003	95.9%	94.0%

Notes: The table shows the percentage of each year's geographic units and annual employment that is in industries covered continuously from 1987 to 2003.

Appendix B: Statistics New Zealand Guide to interpreting the Business Activity Indicator data

[downloaded from:

[http://www2.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Activity+\(GST\)+Indicator#Design](http://www2.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Activity+(GST)+Indicator#Design)]

Guide to Interpreting Data

Usage and Limitations of the Data ..Conceptual differences between GST and other measures

GST sales is not strictly comparable with sales variables measured in other surveys such as Retail Trade Survey or Manufacturing Survey. GST sales and purchases variables are GST inclusive. Statistics New Zealand business surveys record sales and purchases exclusive of GST. The statistical unit for GST is the enterprise, other surveys don't all use enterprises. Some use the kind of activity unit and others the geographic unit.

In addition to the usual sales of goods and services the GST sales variable includes other items such as:

- Sales of second-hand assets. These are normally recorded as capital items in the balance sheet of a business' accounts.
- Sales of businesses themselves. If they are sold as a going concern the sale is zero-rated. The amount of the sale will still appear in the GST sales variable. Some very large sales which breach Statistics New Zealand's confidentiality rules have been removed.

In addition to purchases of goods and services used in the production process the GST purchases variable also includes:

- Purchases of land, buildings, plant and machinery etc, referred to in the National Accounts as Gross Fixed Capital Formation, which is normally recorded in the balance sheet of a business' accounts.
- Purchases of businesses themselves. If the business is sold as a going concern the amount of the sale is not record as a GST purchase. Some very large purchases which breach Statistics New Zealand's confidentiality rules have been removed.

GST sales is not the same as the national accounting concept of gross output.

- Gross output is measured on an accruals basis - businesses have the option of reporting activity on a cash or accruals basis or a combination (see details below).
- Gross output measures sales plus stock change, whereas GST sales is exclusive of stock change.
- Gross output does not record sales of capital goods and services, these appear as Gross Fixed Capital Formation in Expenditure on Gross Domestic Product.
- Gross output does not record sales of businesses.
- Gross output does not include subsidies, whereas GST sales includes any grants or subsidies received.
- Gross output in wholesale and retail industries records the gross margin (sales less cost of goods sold), whereas GST records gross sales.

GST purchases is not the same as the national accounting concept of intermediate consumption.

- Intermediate consumption is measured on an accruals basis - businesses have the option of reporting activity on a cash or accruals basis or a combination (see details below).
- Intermediate consumption measures purchases plus stock change, whereas GST sales is exclusive of stock change.
- Intermediate consumption does not record purchases of capital goods and services, these appear as Gross Fixed Capital Formation in Expenditure on Gross Domestic Product.
- Intermediate consumption does not record purchases of businesses.

Net sales (GST sales less GST purchases) is not the same as the national accounting concept of value added because of the conceptual differences listed above. Value added (GDP) is defined as the value of gross output less the value of intermediate consumption.

Appendix C Matching BD and BAI records (2003)

Appendix Table 3 groups enterprises by their appearance in the BD and BAI datasets. The table reports on the merging of BD records from the February 2003 cross-section, and BAI records for the year ended February 2003.

The first two columns, labelled BD and BAI, indicate whether an enterprise has a BD (BD=1) and/or BAI (BAI=1) record. The third column (FTE) describes the FTE employment size of the enterprise. If an enterprise has a BD record then its number of FTE workers is categorised as either zero (0), greater than zero to two (0-2), or greater than two (.GT.2). The Net GST column categorises an enterprises's GST turnover (sales – purchases) as being equal to zero (0), greater than zero but less than \$40,000 (.LT.40k), or equal to or greater than \$40,000 (.GE.40k).⁴³ The symbol “m” in the FTE and Net GST columns indicates that the group of enterprises have missing FTE employment or net GST. Of the enterprises that are missing either a BD or a BAI record, 258,690 (95.2%) have only a BAI record and the remainder (4.8%) have only a BD record.

Appendix Table 3: Matching enterprises between the BD and BAI datasets in 2003

BD	BAI	FTE	Net GST	Enterprises		FTE	
				#	%	#	%
0	1	m	\$0	47,060	(8.5%)		
0	1	m	.LT.\$40k	86,100	(15.6%)		
0	1	m	.GE.\$40k	59,720	(10.8%)		
0	1	m	m	65,800	(11.9%)		
1	0	(0-2]	m	5,670	(1.0%)	6,500	
1	0	0FTE	m	5,820	(1.1%)		(0.4%)
1	0	.GT.2FTE	m	1,460	(0.3%)	27,840	(1.8%)
1	1	(0-2]	\$0	850	(0.2%)	970	(0.1%)
1	1	(0-2]	.LT.\$40k	56,380	(10.2%)	68,450	(4.5%)
1	1	(0-2]	.GE.\$40k	64,720	(11.7%)	84,760	(5.6%)
1	1	(0-2]	m	20,100	(3.6%)	24,850	(1.6%)
1	1	0FTE	\$0	2,910	(0.5%)		
1	1	0FTE	.LT.\$40k	14,860	(2.7%)		
1	1	0FTE	.GE.\$40k	21,550	(3.9%)		
1	1	0FTE	m	14,750	(2.7%)		
1	1	.GT.2FTE	\$0	290	(0.1%)	3,460	(0.2%)
1	1	.GT.2FTE	.LT.\$40k	7,390	(1.3%)	30,500	(2.0%)
1	1	.GT.2FTE	.GE.\$40k	71,780	(13.0%)	1,197,550	(78.4%)
1	1	.GT.2FTE	m	6,420	(1.2%)	82,310	(5.4%)
					(100%)		(100%)

Notes: Includes all industries (no minimum industry coverage restriction).

Graduated random rounding has been applied to the frequency and cumulative frequency values to protect confidentiality.

⁴³ Note that this is not the same as the threshold for economic significance, which is a threshold for the maximum of sales and purchases.

Appendix D: Labour productivity means tables

Appendix Table 4: Mean (under-represented) productivity by LMA [23/11]

LMA	Ave Lab Prod	LMA	Ave Lab Prod
Wellington	108.4	Thames	35.3
Whangarei	84.6	Te Awamutu	35.1
Balclutha	75.6	Matamata	34.3
Hutt Valley	64.2	Ashburton	33.9
Hawera	63.5	Queenstown	33.7
Auckland	63.3	Bulls	33.4
New Plymouth	60.1	Blenheim	32.9
SthAuckland	59.5	Waihi	32.6
Dunedin	58.7	Hastings	32.5
Te Puke	57.8	Ngaruawahia	32.1
Rotorua	55.0	Dannevirke	31.8
Hamilton	54.5	Dargaville	31.4
Christchurch	48.6	Kaikohe	31.0
Palmerston Nth	45.1	Masterton	30.6
Gore	44.6	Picton	30.5
Taupo	44.6	TeKuiti	30.3
Invercargill	44.1	Taumaranui	28.7
Tokoroa	42.4	Levin	28.3
Morrinsville	42.3	Kaikoura	27.8
Stratford	41.6	Motueka	27.6
Tauranga	38.9	Eketahuna	26.0
Whakatane	38.8	Warkworth	22.0
Otorohanga	38.6	Kerikeri	22.0
Nelson	38.5	MacKenzie	21.8
Alexandra	38.5	Waimate	21.1
Napier	38.0	Kaitaia	20.1
Wanganui	37.4	Taihape	18.1
Waipukurau	36.2	Oamaru	5.9
Gisborne	35.9		
Greymouth	35.3	Total	56.1

Appendix Table 5: Mean (under-represented) productivity by Industry

Industry Group	Average Labour Productivity
D:Electricity, Gas and Water Supply	275.3
M:Government Administration & Defence	177.3
I:Transport and Storage	124.8
F:Wholesale Trade	111.7
B:Mining	110.7
P:Cultural and Recreational Services	56.9
L:Property & Business Services	55.9
C:Manufacturing	51.0
J:Communication Services	44.7
Q:Personal and other Services	37.6
O:Health & Community Services	34.8
K:Finance and Insurance	34.8
E:Construction	34.2
A:Agriculture, Forestry and Fishing	33.3
G:Retail Trade	30.2
N:Education	24.4
H:Accommodation, Cafes & Restaurants	23.2
Total	56.1

Motu Working Paper Series

- 06–07. Grimes, Arthur, David C Maré and Melanie Morten. “Defining Areas Linking Geographic Data in New Zealand”
- 06–06. Maré, David C. and Yun Liang. “Labour Market Outcomes for Young Graduates.”
- 06–05. Hendy, Joanna and Suzi Kerr. “Land-Use Intensity Module: Land Use in Rural New Zealand Version 1.”
- 06–04. Hendy, Joanna, Suzi Kerr and Troy Baisden. “Greenhouse gas emissions charges and credits on agricultural land: what can a model tell us?”
- 06–03. Hall, Viv B., C. John McDermott and James Tremewan. “The Ups and Downs of New Zealand House Prices.”
- 06–02. McKenzie, David; John Gibson and Steven Stillman, “How Important is Selection? Experimental Vs Non-Experimental Measures of the Income Gains from Migration.”
- 06–01. Grimes, Arthur and Andrew Aitken, “Housing Supply and Price Adjustment.”
- 05–14. Timmins, Jason, “Is Infrastructure Productive? Evaluating the Effects of Specific Infrastructure Projects on Firm Productivity within New Zealand.”
- 05–13. Coleman, Andrew; Sylvia Dixon and David C. Maré, “Māori Economic Development – Glimpses from Statistical Sources.”
- 05–12. Maré, David C., “Concentration, Specialisation and Agglomeration of Firms in New Zealand.”
- 05–11. Holmes, Mark J. and Arthur Grimes, “Is there Long-run Convergence of Regional House Prices in the UK?”
- 05–10. Hendy, Joanna and Suzi Kerr, “Greenhouse Gas Emission Factor Module: Land Use in Rural New Zealand–Climate Version 1.”
- 05–09. Poland, Michelle and David C. Maré, “Defining Geographic Communities.”
- 05–08. Kerr, Suzi; Joanna Hendy, Emma Brunton and Isabelle Sin, “The Likely Regional Impacts of an Agricultural Emissions Policy in New Zealand: Preliminary Analysis.”
- 05–07. Stillman, Steven, “Examining Changes in the Value of Rural Land in New Zealand between 1989 and 2003.”
- 05–06. Dixon, Sylvia and David C. Maré, “Changes in the Māori Income Distribution: Evidence from the Population Census.”
- 05–05. Sin, Isabelle and Steven Stillman, “The Geographical Mobility of Māori in New Zealand.”
- 05–04. Grimes, Arthur, “Regional and Industry Cycles in Australasia: Implications for a Common Currency.”
- 05–03. Grimes, Arthur, “Intra and Inter-Regional Industry Shocks: A New Metric with an Application to Australasian Currency Union.”
- 05–02. Grimes, Arthur; Robert Sourell and Andrew Aitken, “Regional Variation in Rental Costs for Larger Households.”
- 05–01. Maré, David C., “Indirect Effects of Active Labour Market Policies.”
- 04–12. Dixon, Sylvia and David C Maré, “Understanding Changes in Maori Incomes and Income Inequality 1997–2003.”
- 04–11. Grimes, Arthur, “New Zealand: A Typical Australasian Economy?”
- 04–10. Hall, Viv and C. John McDermott, “Regional Business Cycles in New Zealand: Do They Exist? What Might Drive Them?”
- 04–09. Grimes, Arthur; Suzi Kerr and Andrew Aitken, “Bi-Directional Impacts of Economic, Social and Environmental Changes and the New Zealand Housing Market.”

- 04-08. Grimes, Arthur and Andrew Aitken, "What's the Beef with House Prices? Economic Shocks and Local Housing Markets."
- 04-07. McMillan, John, "Quantifying Creative Destruction: Entrepreneurship and Productivity in New Zealand."
- 04-06. Maré, David C. and Isabelle Sin, "Maori Incomes: Investigating Differences Between Iwi."
- 04-05. Kerr, Suzi; Emma Brunton and Ralph Chapman, "Policy to Encourage Carbon Sequestration in Plantation Forests."
- 04-04. Maré, David C., "What do Endogenous Growth Models Contribute?"
- 04-03. Kerr, Suzi; Joanna Hendy, Shuguang Liu and Alexander S.P. Pfaff, "Uncertainty and Carbon Policy Integrity."
- 04-02. Grimes, Arthur; Andrew Aitken and Suzi Kerr, "House Price Efficiency: Expectations, Sales, Symmetry."
- 04-01. Kerr, Suzi; Andrew Aitken and Arthur Grimes, "Land Taxes and Revenue Needs as Communities Grow and Decline: Evidence from New Zealand."
- 03-19. Maré, David C., "Ideas for Growth?"
- 03-18. Fabling, Richard and Arthur Grimes, "Insolvency and Economic Development: Regional Variation and Adjustment."
- 03-17. Kerr, Suzi; Susana Cardenas and Joanna Hendy, "Migration and the Environment in the Galapagos: An Analysis of Economic and Policy Incentives Driving Migration, Potential Impacts from Migration Control, and Potential Policies to Reduce Migration Pressure."
- 03-16. Hyslop, Dean R. and David C. Maré, "Understanding New Zealand's Changing Income Distribution 1983-98: A Semiparametric Analysis."
- 03-15. Kerr, Suzi, "Indigenous Forests and Forest Sink Policy in New Zealand."
- 03-14. Hall, Viv and Angela Huang, "Would Adopting the US Dollar Have Led to Improved Inflation, Output and Trade Balances for New Zealand in the 1990s?"
- 03-13. Ballantyne, Suzie; Simon Chapple, David C. Maré and Jason Timmins, "Movement into and out of Child Poverty in New Zealand: Results from the Linked Income Supplement."
- 03-12. Kerr, Suzi, "Efficient Contracts for Carbon Credits from Reforestation Projects."
- 03-11. Lattimore, Ralph, "Long Run Trends in New Zealand Industry Assistance."
- 03-10. Grimes, Arthur, "Economic Growth and the Size & Structure of Government: Implications for New Zealand."
- 03-09. Grimes, Arthur; Suzi Kerr and Andrew Aitken, "Housing and Economic Adjustment."
- 03-07. Maré, David C. and Jason Timmins, "Moving to Jobs."
- 03-06. Kerr, Suzi; Shuguang Liu, Alexander S. P. Pfaff and R. Flint Hughes, "Carbon Dynamics and Land-Use Choices: Building a Regional-Scale Multidisciplinary Model."
- 03-05. Kerr, Suzi, "Motu, Excellence in Economic Research and the Challenges of 'Human Dimensions' Research."
- 03-04. Kerr, Suzi and Catherine Leining, "Joint Implementation in Climate Change Policy."
- 03-03. Gibson, John, "Do Lower Expected Wage Benefits Explain Ethnic Gaps in Job-Related Training? Evidence from New Zealand."
- 03-02. Kerr, Suzi; Richard G. Newell and James N. Sanchirico, "Evaluating the New Zealand Individual Transferable Quota Market for Fisheries Management."
- 03-01. Kerr, Suzi, "Allocating Risks in a Domestic Greenhouse Gas Trading System."

All papers are available online at http://www.motu.org.nz/motu_wp_series.htm