GEOGRAPHIC CONCENTRATION OF NEW ZEALAND EMPLOYMENT[1,2]

David C Maré

Motu Economic and Public Policy Research Trust

Abstract

This paper examines the degree of geographic concentration of employment in New Zealand, using summary measures proposed by Ellison and Glaeser (1997) and Maurel and Sedillot (1999). We use Statistics New Zealand Business Demography data for the period 1987-2003, and find that concentration has risen during that period. Around half of employment is in industries that display no significant concentration. International comparisons of the degree of concentration are difficult, but it appears that New Zealand's levels are similar to those of the United Kingdom, and lower than those of the US and France. Where concentration does occur, it operates most strongly over distances of less than 50km.

Introduction

Current policy interest in 'sustainable cities', regional development, cluster development, and 'Regional Centres of Excellence' reflects an acceptance that local factors matter for firm performance. It is plausible that location is important for growth, innovation, and productivity, although determining how and why it matters is far more difficult. The existence of cities, and of geographically uneven patterns of activity can be seen as *prima facie* evidence that there are advantages to locating where others locate. In a recent survey of relevant theories of economic geography and agglomeration, Ottaviano and Thisse (2003) note that "peaks and troughs in the spatial distributions of population, employment and wealth are a universal phenomenon in search of a general theory".

The lack of such a general theory is particularly problematic for policy-makers, whose attempts to reinforce the benefits of agglomeration will be hampered by an inability to identify the nature of those benefits. In a separate review of agglomeration theories, with a particular emphasis on cities, Duranton and Puga (2003), summarise the challenges as follow:

"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]

Rosenthal and Strange (2003) and Crawford (2004) summarise the empirical literature on the nature, sources and effects of agglomeration economies, which is similarly inconclusive. Both surveys document evidence for a wide range of potential mechanisms, but both point to the difficulties of linking empirical studies to specific agglomeration mechanisms.

Measuring Geographic Concentration

The most commonly used indices in the recent economics literature are the area-based indices developed by Ellison and Glaeser (1994) (EG) and Maurel and Sedillot (1999) (MS). An alternative to the area-based measures is a distance-based measure such as that proposed by Duranton and Overman (2002). This section introduces the measures that will be used in the remainder of the paper.

Area-based Measures of Concentration

Both Maurel and Sedillot (1999) (MS) and Ellison and Glaeser (1994) (EG) have proposed indices that arise from a statistical model as estimators of the correlation between location decisions of two firms. The presentation that follows is based closely on Maurel and Sedillot (1999).

In the MS model, there are N plants in an industry, with industry employment shares $z_1 \ldots z_N$. These are located across M geographic areas, and $x_1 \ldots x_M$ are the area shares of total employment. The fraction of the industry's employment that is in area i is

$$s_i = \sum_{j=1}^{N} z_j u_{ji} \tag{1}$$

where $u_{ji} = 1$ if firm j locates in area i, and zero otherwise. The u_{ji} are non-independent binomial variables for which $P(u_{ji} = 1) = x_i$ so that firm location decisions are expected to aggregate to the observed distribution of total employment. The MS index of concentration (γ) is an estimate of the correlation of u_{ji} and u_{ki} for two firms, j and k. The probability that two firms j and k locate in the same area (i) is:

$$P(i,i) = E\left(u_{ji}, u_{ki}\right)$$
$$= Cov\left(u_{ji}, u_{ki}\right) + E\left(u_{ji}\right)E\left(u_{ki}\right) \qquad (2)$$
$$= \gamma x_i (1 - x_i) + x_i^2$$

which suggests an estimator for γ of:

$$\hat{\gamma} = \frac{\hat{p} - \sum_i x_i^2}{1 - \sum_i x_i^2} \qquad (3)$$

Maurel and Sedillot (1999) suggest a natural frequency-based estimator for $\hat{p}$, the probability that two firms from within the same industry locate together: (The Appendix to Maurel and Sedillot (1999) contains a more detailed derivation.)

$$\hat{p} = \frac{\sum_{\substack{i,k \in i \\ j \neq k}} z_j z_k}{\sum_{\substack{i,k \\ j \neq k}} z_j z_k} = \frac{\sum_i s_i^2 - \sum_i z_i^2}{1 - \sum_i z_i^2} \qquad (4)$$

$$= \frac{\sum_i s_i^2 - H}{1 - H}$$

where H is the familiar Herfindahl index of industrial (employment) concentration, being the sum of squared plant shares in industry employment. Substituting equation (4) into (3) yields the MS concentration index:

$$\hat{\gamma} = \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2} - H}{1 - H} \qquad (5)$$

$$= \frac{G - H}{1 - H}$$

The adjustment for the Herfindahl index has a natural interpretation. An industry with a single plant will necessarily be located in a single location, even if the choice of location were totally random. We do not want to classify an industry as concentrated just because employment is concentrated in a small number of plants.

The MS index is closely related to the earlier EG index, the only difference being in the form of G, which Ellison and Glaeser (1994) derive based on an *a priori* reasoning. The formula for the EG index is

$$\hat{\gamma}_{EG} = \frac{\dfrac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^2} - H}{1 - H} = \frac{G_{EG} - H}{1 - H} \qquad (6)$$

Both indices are unbiased estimators of the correlation of u_{ji} and u_{ki}. The slight difference in formula can, however, lead to quite different inferences about whether an industry is geographically concentrated or not.

The difference between the two indices reflects the difference between the terms $(s_i - x_i)^2$ in the numerator of

G_{EG} and $(s_i^2 - x_i^2)$ from the numerator of G. The difference is $x_i(s_i - x_i)$, which is positive when the industry is over-represented in areas where total employment is concentrated, and negative when it is over-represented where the total employment share is small.

Area-based Measures of Co-location

It is relatively straightforward to adapt the area-based measures of industry concentration to measure the degree to which different industries have common locational patterns.

This index is derived by applying the Maurel and Sedillot framework to provide an estimate of the correlation of location decisions between two firms from different industries (rather than between two firms within the same industry). With this change, equation (4) now takes the following form:

$$\hat{p} = \frac{\sum_{j,k \in i} z_j z_k}{\sum_{j,k} z_j z_k} = \sum_i s_i^m s_i^n \qquad (7)$$

where firm j belongs to industry m, firm k belongs to industry n, and s_i^w is the share of industry w employment that is in area i. This implies a a co-location index γ_{co} of

$$\gamma_{co}(m,n) = \frac{\sum_i s_i^m s_i^n - \sum_i x_i^2}{1 - \sum_i x_i^2} \qquad (8)$$

This index attains a value of zero when both industries are located proportionately to total employment, is highest when the two industries are located together in a single location, and is lowest (most negative) when the two industries are each fairly evenly spread across locations but never locate in the same location.

Distance-based Measures

An alternative to using area-based measures of concentration as outlined above is to use distance between plants or jobs as the basis of a continuous measure. We adopt the general approach of Duranton and Overman (2002), which analyses the distribution of bilateral distances between jobs or plants. More specifically, we use their measures of 'local localisation and dispersion', with plant locations weighted by employment.

Using Duranton and Overman's notation, let e_i and e_j be the number of jobs at location i and j respectively, and let $D(i,j)$ be the straight-line distance between location i and location j. If there are n different locations, there are $n(n-1)/2$ unique bilateral distances. We calculate a frequency distribution for these distances, and summarise the density.

Define an indicator variable $\delta(i,j,d)$ that takes the value one when $D(i,j)=d$, and zero otherwise. The frequency at distance d is:

$$K(d)=\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\delta(i,j,d)e(i)e(j)+\sum_{i=1}^{n}\frac{\delta(i,i,d)e(i)(e(i)-1)}{2}}{\sum_{i=1}^{n}\sum_{j=i}^{n}e(i)e(j)+\sum_{i=1}^{n}\frac{e(i)(e(i)-1)}{2}} \qquad (9)$$

This differs from Duranton and Overman's equation 2 because, unlike Duranton and Overman, we include zero distances. They exclude from their frequency distribution instances where $i=j$ because they do not want their summary measure to be influenced by large plants. Unfortunately, we do not observe the exact locations of plants, although we do observe location at a fairly detailed (meshblock) level. A number of plants can thus be observed in the same location and we wish to count distances between nearby plants in the distance density, even though they are observed at zero distance.

The resulting frequency distributions are calculated for total employment as well as for each 2-digit and 4-digit industry, and for groupings of industries as described in the following section.

Kernel Density Smoothing

Each frequency distribution $K(d)$ is smoothed using kernel density methods. We use the same smoothing method as used by Duranton and Overman (2002), which makes use of a Gaussian kernel, with bandwidth set by Silverman's rule of thumb.

Bootstrap Standard Errors

Bootstrap standard errors are estimated from the 'total employment' distance density. In order to judge whether a specific industry's distance density differs significantly from the total density, we sample with replacement from the total population of jobs, and calculate a distance density from the selected sample. The size of the sample is chosen to equal the number of jobs in the industry.

In practice, the derived bootstrap standard errors were approximate. We carried out 50 replications, each with a sample of 50,000. Approximate 95 percent confidence intervals at each distance were estimated as 1.96 times the standard deviation of the densities at that distance. This single set of distance-specific standard errors was then scaled for each industry, to reflect the employment in that industry. For an industry with employment E, the scaling factor was $\sqrt{50,000}/\sqrt{E}$. This approach gave confidence intervals that were extremely close to confidence intervals using 250 replications and separately drawn replications, at least for a selection of 2-digit industries. Our approach thus yields bands that are similar to Duranton and Overman's 'local confidence' bands.

Data

The data used in this study are taken from Statistics New Zealand's (SNZ) business demography datasets, and provide annual longitudinal data on the majority of New Zealand businesses from 1987 to 2003, measured as at February each year. 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 target population for these 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 SNZ Business Frame at that point in time. From 1987 to 1994, the data are taken from the SNZ Business Directory, and from 1994 to 2003, they are from the SNZ Business Frame.

The data are collected from a combination of survey and administrative sources – primarily the SNZ Annual Business Frame Update Survey (ABFU) 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.

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. In this paper, we deal exclusively with business units and not enterprises.

The criteria for including activity units in the database is described in detail in Carroll et al. (2002) and Statistics New Zealand (2004). A major change in the data is the shift from GST-registration to economic significance, which occurred in 1994. From 1987 to 1994, business units were included only if they belonged to a GST-registered enterprise (i.e. with GST sales of at least $30,000). From 1994 the business unit was included only if it belonged to an 'economically significant' enterprise, where an enterprise was regarded as economically significant if it met any one of the following criteria: had greater than $30,000 annual GST expenses or sales; had more than 2 full-time equivalent paid employees; was in a GST-exempt industry except residential property leasing and rental; or; was part of a group of enterprises.

To gauge the impact of the series discontinuity in 1994 on our analysis of changes over time, we examined the time pattern of employment and the number of business units – in total and disaggregated by industry, firm-size, and region. In aggregate, the change in definition caused a 10 percent decline in the number of business units covered, and a 1 percent decline in measured FTE employment. As would be expected, the changes were more noticeable across the firm-size distribution, with the declines resulting from definitional change being confined to small (0-5 FTE) firms. Given that most of our analyses are employment-weighted, the impact is likely to be small. Furthermore, it appears that the geographic and industry impact of the changes were widespread, so that the

discontinuities in the sort of indices that we are looking at will be minor. While we remain cautious in interpreting any changes around 1994, we consider that pooled analysis is still justified.

The industry coverage of the business demography data has also changed over time. The primary exclusion from the BDS is firms in agricultural production industries. 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. The analysis in this paper is restricted to 4-digit industries that remained in coverage throughout the study period. Apart from the exclusion of agricultural industries that were surveyed in 1998, this restriction is most severe in the most recent period, where excluded industries contained just over 4 percent of full-time-equivalent employment (FTE = FT + 0.5*PT, including working proprietors).

The locations of business units are recorded at the level of meshblocks, which provide a high level of geographical detail. For the distance-based analyses, we use 1991 meshblocks, with the location of each meshblock fixed at its geographic centroid. Around 75 percent of meshblocks lie entirely within 1 kilometre of their centroid. Over 91% of the population lives in meshblocks where their true location is within 1 kilometre of the centroid.

Our main analyses group employment into 58 labour market areas (LMAs), as defined by Newell and Papps (2001) on the basis of commuting patterns. Newell and Papps (2001) define two sets of labour market areas – one with 140 areas and one with 58. We have chosen to use the more aggregated areas.

Analysis Subsamples

Of the 424 4-digit industries that are represented in the data in at least one period, 17 are dropped because they are out of coverage in some periods. Of the remaining 407 industries, information is suppressed for some industries, in order to protect confidentiality. In order to examine trends over time, we exclude industries that have suppressed information in *any* year. Ninety-four 4-digit industries are dropped due to suppression, leaving 323 4-digit industries.

When examining the overall prevalence of concentration, we calculate indices for each industry for each year and then average each industry across all years where information is not suppressed. Information can be used from industries that are suppressed in some but not all years. Thirty-four 4-digit industries are dropped due to suppression, leaving .390 industries.

Co-location and distance-based analyses are undertaken using only 2003 data, although still with minimum industry coverage imposed. Fifty 4-digit industries are dropped from the co-location analysis due to suppression,

leaving 374. For the distance-based analysis, a further 20 industries are dropped because of the information contained in distance densities, leaving 354 usable industries.

Trends in Concentration

Figure 1 shows the trends in the geographic concentration of New Zealand 4-digit industries across LMAs, for the 1988 to 2003 period. The measures shown are the Maurel-Sedillot index (MS), and the Ellison-Glaeser index (EG). Also shown is the industry Herfindahl, which captures the degree to which industry employment is dominated by a few business sites. All series are employment-weighted averages of industry-specific measures. The sample is restricted to 323 industries that were in the sample throughout the period, and for which no data were suppressed in any year.

There has been a clear increase in the degree of concentration over time, with a particularly pronounced rise in the early 1990s. The increases were widespread, with relatively large contributions from manufacturing, resource-based, and service industries. In analysis not shown here, we find that employment growth has not been significantly greater for concentrated industries than for less concentrated industries. The trend increase in concentration is thus due to industries, on average, increasing their degree of concentration, and not to compositional change.

The decline in the average Herfindahl index was also widespread, although with strong contributions from central government services and transport industries in the late 1980s.

Interestingly, the rise in geographic concentration does not occur for smaller geographic areas. When observing location at the level of area units rather than LMAs, concentration declines over the study period. It would appear that, although industries are becoming more concentrated in particular LMA (commuting zones), their dispersion within each commuting zone is, on average, becoming greater.

How Prevalent is Concentration?

To help in summarising the prevalence of concentration across 4-digit industries, we group together industries that have similar locational patterns. We do this based on the co-location index (γ_{co}) described earlier. Collecting together these indices for each pair of industries, we can form a correlation matrix, which summarises the strength of co-location effects. We first transform the correlation matrix into a matrix of dissimilarity or distance measures by subtracting each entry from 1. Using this distance matrix, we apply a standard statistical clustering procedure to perform a hierarchical clustering of industries.

Figure 1: Trends in concentration – 4-digit industries across Labour Market Areas (LMA)

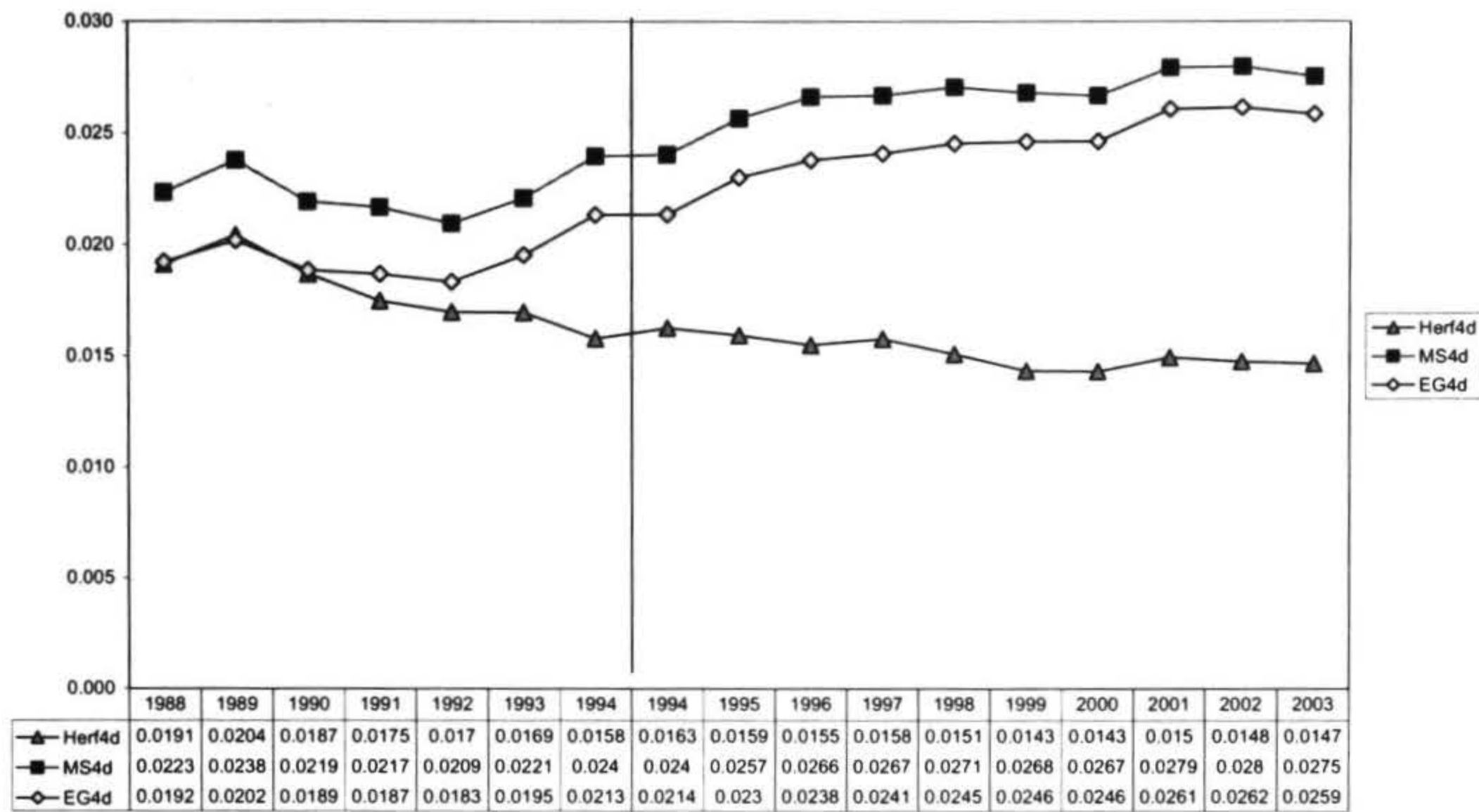|	1988	1989	1990	1991	1992	1993	1994	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Herf4d	0.0191	0.0204	0.0187	0.0175	0.017	0.0169	0.0158	0.0163	0.0159	0.0155	0.0158	0.0151	0.0143	0.0143	0.015	0.0148	0.0147
MS4d	0.0223	0.0238	0.0219	0.0217	0.0209	0.0221	0.024	0.024	0.0257	0.0266	0.0267	0.0271	0.0268	0.0267	0.0279	0.028	0.0275
EG4d	0.0192	0.0202	0.0189	0.0187	0.0183	0.0195	0.0213	0.0214	0.023	0.0238	0.0241	0.0245	0.0246	0.0246	0.0261	0.0262	0.0259

Table 1 lists the resulting groups, showing the number of industries in each group, the number of geographic units, employment shares, and employment weighted averages for Herfindahl, MS and EG indices. The groups are listed in descending order of concentration, as measured by the MS index. The titles given to the groups are indicative, and while providing a useful shorthand, they do not convey the broad range of industries that are included in each group.

The number of groups is chosen by the researcher. I have chosen eight groups, which fairly reflects the main patterns in the data. Roughly two-thirds of employment is in industry groups that do not display significant geographic concentration. Slightly more than a quarter of employment is in geographically concentrated industry groups. Resource-based industries, which have a high EG index but which are very dispersed according to the MS index, account for most of the remainder.

The largest group is the 'Local Services' group, which has concentration measures close to zero, and accounts for almost half of all employment. The industries generally provide local goods and services, and are distributed fairly well in proportion to total employment. It seems likely that proximity to output markets is an important factor in location decisions for these industries, although there may be other interactions as well. A further 15% of employment is in another group of low-concentration industries grouped under the heading of 'Dispersed Manufacturing'. Proximity to local markets seems to be a plausible explanation for the location patterns of industries in this group as well.

Geographic concentration is high for four groups that together account for 27.9 percent of employment. The most concentrated group is the 'Centralised Services' group of 7 industries, which is dominated by central government and accounts for 2.7% of FTE employment. The 'Business Services and Consumer Wholesaling' group is the largest of the concentrated groups, accounting for 18 percent of employment. This is the group that, (speculatively at least) is most likely to be influenced by agglomeration forces such as knowledge spillovers.

The 'Concentrated Manufacturing and Wholesaling' group adds a further 7.1 percent of employment. The industries in this group tend to be producers of heavier goods that may benefit from proximity to shared (transport) infrastructure or concentrated input suppliers, or industries for which economies of scale are important.

The very small 'Chemical and Mining' group of 5 industries also appears as concentrated, although it accounts for only 0.1 percent of employment.

Just over 7 percent of employment is in the 'Resource-based' group of industries, which have a negative MS index and high levels of concentration according to the EG index. As noted already these are industries in which employment is distributed more evenly across LMAs than is total employment, and is disproportionately in LMAs where the share of total employment is high. Common reliance on land and land-based resources would appear to be an important cause of agglomeration for these industries.

Table 1: Co-location groupings

Employment weighted	# Inds	Units	FTE	FTE %	Herf	MS	EG
Centralised services	5	1,600	37,600	2.7	0.020	0.131	0.132
Business Services and Consumer wholesaling	78	60,800	264,800	19.0	0.013	0.121	0.043
Concentrated Manufacturing & Wholesaling	64	12,200	104,275	7.5	0.052	0.120	0.074
Chemicals and Mining	4	50	1,130	0.1	0.310	0.035	0.072
Dispersed Manufacturing	68	37,100	222,660	16.0	0.018	0.001	0.010
Local services	105	124,600	644,850	46.3	0.009	-0.008	0.001
High Herfindahl industries	10	570	11,115	0.8	0.156	-0.012	0.024
Resource-based	40	13,500	105,115	7.6	0.029	-0.037	0.095
Total	374	250,300	1.39m				

Note: Herfindahl, MS, and EG indices are employment-weighted averages of values for each 4-digit industry.

How Does New Zealand Compare Internationally?

One obvious question that arises in interpreting concentration indices is whether the degree of concentration in New Zealand is higher or lower than in other OECD countries.

International comparisons of area-based indices such as the MS and EG indices are, at best, imperfect, given the impossibility of using directly comparable area definitions. Given New Zealand's relatively low geographic density of economic activity, it is not possible to match both on the size of geographic areas and on employment numbers. A number of studies have, nevertheless, compared indices from studies of different countries. The most common approach is to compare unweighted averages of indices for 4-digit industries. This is the statistic that is available from the greatest range of studies, including those for the US, UK, and France, as presented in Ellison and Glaeser (1994) Devereux et al. (2004) and Maurel and Sedillot (1999) respectively.

The means and medians of the various concentration indices are shown in Table 2, for total employment and for manufacturing alone. The first panel of the table shows indices weighted by employment, whereas the second panel shows them unweighted. The unweighted measures calculated for 4-digit manufacturing industries provide a basis for comparing New Zealand's geographic concentration with that of other countries.

The employment weighting generally lowers estimated concentration, reflecting the fact that larger industries tend to have lower concentration. The exception to this is for the concentration of 4-digit manufacturing industries across LMAs. The reason appears to be that there are a number of relatively large primary-sector-related manufacturing industries that have high EG indices (and negative MS indices). Weighting by employment emphasises these industries, and raises the average. It also magnifies the difference between the MS and EG indices. The sensitivity of the New Zealand measures to weighting makes it more difficult to draw firm conclusions about whether New Zealand employment is more or less concentrated than that of other countries.

The employment-weighted average indices provide the best basis for comparison, as they are not as sensitive to extreme values for small industries. Weighting does not appear to be as important a consideration for countries other than New Zealand. Devereux et al. (2004) calculate both MS and EG indices, and find that the two indices are almost identical. This is clearly not the case in New Zealand. Similarly, Table 2 shows the employment-weighted average value of the EG index for the US, calculated from the data provided in Ellison and Glaeser (1994). The weighted EG for United States manufacturing is 0.045, as compared with 0.051 unweighted. On balance, we consider that the weighted index is the more appropriate measure for New Zealand, and that it is reasonable to compare this measure with unweighted measures from other countries.

Overall, although the comparisons are cloudy, it appears that New Zealand has lower concentration than France and the US, and similar to that of the UK. The weighted New Zealand EG measure (0.030) is closer to that of the unweighted UK measure (0.033) than to that of the US (0.051). Similarly, the weighted New Zealand MS measure (0.039) is closer to the UK measure (0.033) than to that of France (0.06). Using the inferior unweighted measures, New Zealand's manufacturing MS index of 0.067 appears high relative to that of France and the UK, and New Zealand's unweighted EG index appears to be anomalously low (0.013).

Table 2: Summary of Concentration Indices

Industry	Area	Herfindahl Index	Maurel-Sedillot Index	Ellison Glaeser Index	Gini
Weighted					
New Zealand					
4-digit (390)	LMA (58)	0.022 [0.006]	0.028 [0.007]	0.025 [0.008]	0.400 [0.377]
4-digit Manuf (137)	LMA (58)	0.057 [0.032]	0.039 [0.025]	0.030 [0.018]	0.571 [0.585]
United States					
4-digit Manuf (459)	State (51)	0.017 [0.008]		0.045 [0.023]	
Unweighted					
New Zealand					
4-digit (390)	LMA (58)	0.062 [0.028]	0.059 [0.026]	0.034 [0.015]	0.488 [0.495]
4-digit Manuf (137)	LMA (58)	0.108 [0.077]	**0.067** [**0.056**]	**0.013** [**0.010**]	0.578 [0.582]
United States					
4-digit Manuf (459)	State (51)	0.028 [0.016]		0.051 [0.026]	
France					
4-digit Manuf (273)	Dept (95)		0.06 [0.01]		
United Kingdom					
4-digit Manuf (211)	Postcode (113)		0.033 [0.006]	0.033 [0.007]	

Notes: The NZ industry-level measures are derived by averaging annual estimates from all years where data were not suppressed for confidentiality reasons. Weighted measures are weighted by average FTE employment. Industry coverage for all NZ measures is restricted to those industries that were covered in the SNZ Business Demography data throughout the period.

The Geographic Scope of Concentration

The analysis so far has been based on boundaried areas – either LMAs or Area Units. Although using commuting-based LMAs reduces the arbitrariness of the boundaries, the indices draw no distinction between concentrations in two contiguous areas and two areas that are distant from each other. Concentration in contiguous arguably provides greater evidence of agglomeration. An alternative approach is to measure concentration on the basis of geographic distance, as was done by Duranton and Overman (2002). As noted earlier, the essence of the approach is to measure the bilateral distance between each pair of jobs and then examine the proportion of pairs that occur at each distance. Distances range between zero and 1468km, with a mean distance of 405km and a median distance of 360km.

In order to investigate the variation in distance densities between industries and industry groups, we focus our attention on distances between 0 and 150 km, and we smooth the densities using kernel density smoothing. We focus attention on a limited geographic range because of our interest in interaction-related agglomeration, and in patterns of concentration within rather than between settled areas. Duranton and Overman (2002) focus on a range of 180 km, which is the median distance in their dataset. Our choice of range is more restrictive, as only 27 percent of bilateral distances are between zero and 150km.

Figure 2 shows the distance density for the highly concentrated 'Business Services and Consumer Wholesaling' industry group. The figure contains four series. The two dotted lines indicate the upper and lower bounds of the 95 percent confidence interval for the aggregate density. The bounds are unweighted averages of the bounds for each industry, averaged at each distance. The lighter solid line shows the unweighted average of *within-industry* densities for industries in the group. The darker solid line shows the employment-weighted average density. Each industry's density is weighted by the size of its FTE employment. The interpretation of this line is that it shows the proportion of *within-industry* job pairs that occur, on average, at each distance. The employment-weighted density is less concentrated at short distances than the unweighted, reflecting the fact that smaller industries tend to be more geographically concentrated.

The density shows significant localisation between 0km and 50km, with a peak of localisation at around 10km. The density is similar for the other two concentrated groups – Concentrated Manufacturing and Wholesaling, and Centralised Services, although the range of localisation for the centralised services group is narrower, between 0 and 20km.

In contrast, the Dispersed Manufacturing group, and the Local Services group have distance densities that are very similar to the total density, although there are ranges for which the group density lies outside the confidence bounds of the total density. Dispersed Manufacturing industries show localisation below 20km and between 90km and 110km, but these peaks are greatly reduced when industries are weighted by employment.

The density for the group of Resource-based industries is shown in Figure 3. This group has a distinct distance density, reflecting relatively low geographic concentration. It is much flatter than the pooled density for all industries. Employment in this group is *less* localised (more dispersed) than average below 30km, and is localised in the 40km to 90km range.

Figure 2: Distance density for 'Business Services and Consumer Wholesaling' industry group
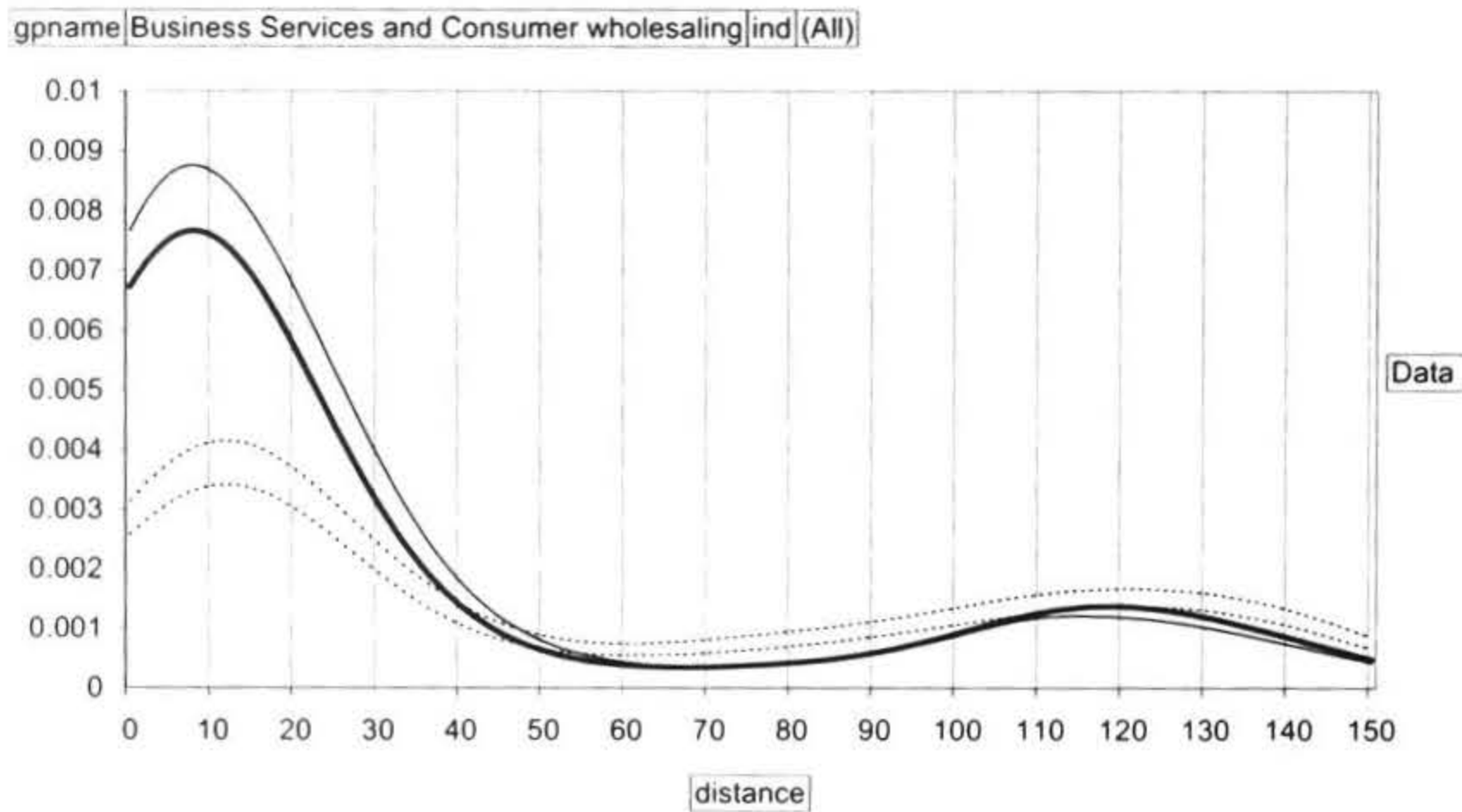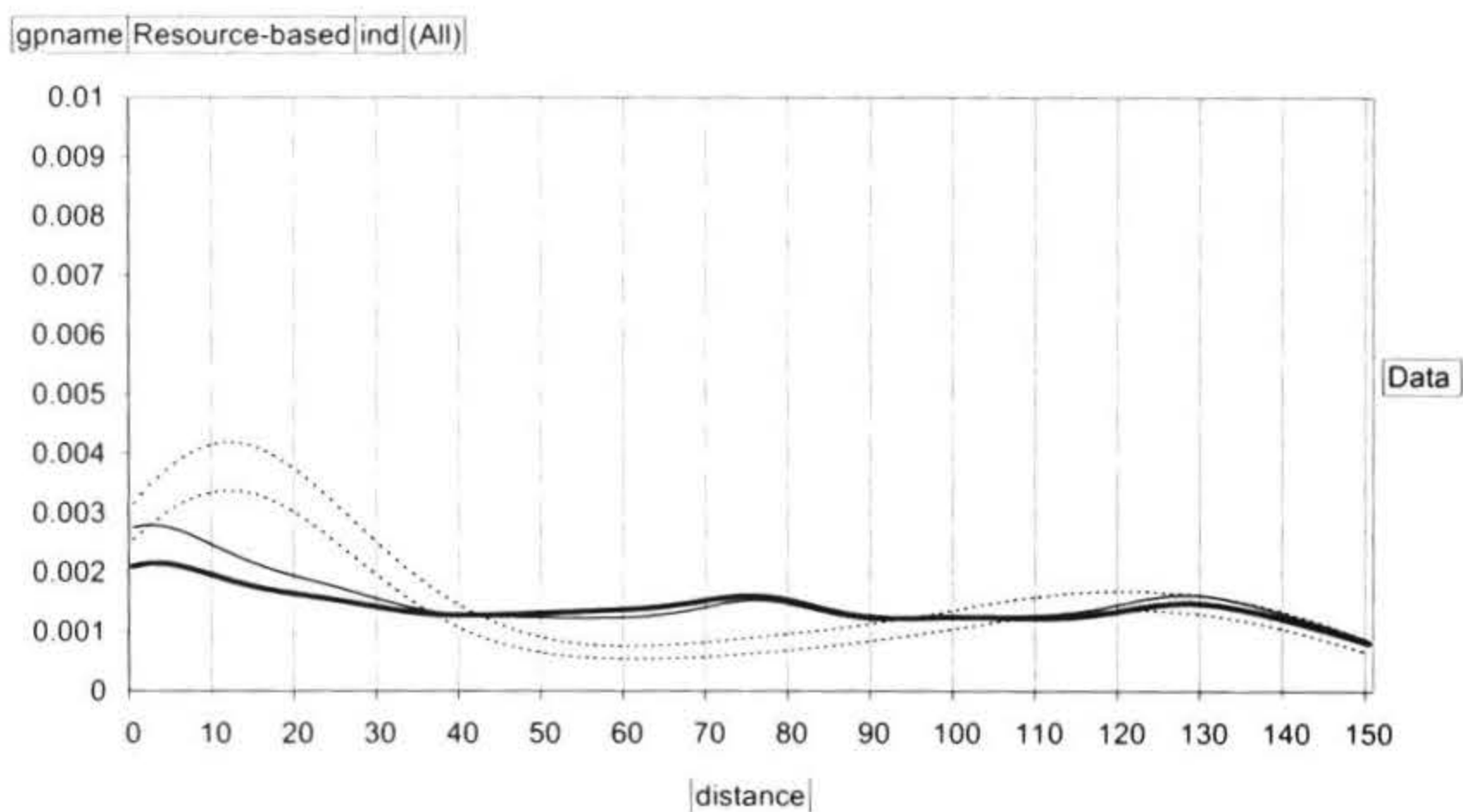

Figure 3: Distance density for 'Resource-based' industry group

Conclusions

Geographic concentration of employment is an obvious feature of modern labour markets. In this paper we have summarised the extent of concentration in New Zealand, looked at changes over time, and investigated the geographic scope of concentration.

Concentration has been rising over the 1987-2003 period of our study, and is currently at levels similar to those found in the United Kingdom, and lower than those of the United States and France.

Four-digit industries were allocated to groups that shared similar locational patterns. Around thirty percent of employment is in industries that show significant concentration. These industries tend to be in business services, wholesaling, and some parts of manufacturing, as well as centralised services (which is dominated by central government). Around half of employment is in local service or dispersed manufacturing industries that are located in proportion to total employment.

Concentrated industries are significantly localised at relatively short distances – between zero and 50 kilometres.

Future Directions

The next stage in this research is to look at the links between concentration and economic performance. Initially, we will look at links with firm survival and with firm and job turnover. We will subsequently investigate whether geographic concentration leads to higher productivity levels or growth, using net GST payments as a proxy for value added.

Another valuable extension would be to examine concentration patterns and trends for particular subsets of industries – eg: those that are involved in importing and exporting; those that invest most in research and development; or those that share a common science base. All of these subsets have been examined in the international literature, and suggest particular types of agglomeration benefits.

A more detailed and extended version of this paper is forthcoming as a Motu Working Paper, which will be downloadable from www.motu.org.nz.

Notes

1. The results presented in this study are the work of the author, not Statistics New Zealand, and do not necessarily reflect the views of the Ministry of Economic Development.

2. This paper has been produced with funding from the Ministry of Economic Development. My thanks to Ron Crawford for helpful and stimulating discussions about the issues covered in the paper, and to participants in previous seminars at the Ministry of Economic Development and the NZ Association of Economists.

References

Carroll, N., Hyslop, D. R., Maré, D. C., Timmins, J. and Wood, J. (2002) An Analysis of New Zealand's Business Demography Database. *Paper presented to the 'Data Integration and Linked Employer-Employee Data' Conference, Wellington,* **21-22 March 2002**.

Crawford, R. (2004) The effects of agglomeration on economic activity - the empirical evidence on mechanisms and magnitudes. *Draft Report,* Wellington: Ministry of Economic Development.

Devereux, M. P., Griffith, R. and Simpson, H. (2004) The geographical distribution of production activity in the UK. *Regional Science and Urban Economics,* **34,** 533-64.

Duranton, G. and Overman, H. (2002) Testing for localisation using micro-geographic data. *Centre for Economic Policy Research Discussion paper,* **3379.**

Duranton, G. and Puga, D. (2003) Microfoundations of urban agglomeration economies. In J. V. Henderson and J. F. Thisse (Eds.), *(Draft for) Handbook of Urban and Regional Economics, Volume 4,* http://econ.pstc.brown.edu/faculty /henderson/WillAndStuart.pdf or http://www. core.ucl.ac.be/staff/thisseHandbook/Handbook.h tml.

Ellison, G. and Glaeser, E. L. (1994) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *NBER Working Paper No. w4840.*

--- (1997) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy, Vol. 105, no. 5 (1997): 889-927.,* **105 (5),** 889-927.

Maurel, F. and Sedillot, B. (1999) A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics,* **29(5),** 575-604.

Newell, J. O. and Papps, K. (2001) Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study using Travel-to-Work Data. *Department of Labour Occasional Paper,* **2001/6.**

Ottaviano, G. and Thisse, J.-F. (2003) Agglomeration and Economic Geography. In J. V. Henderson and J.-F. Thisse (Eds.), *(Draft for) Handbook of Urban and Regional Economics, Volume 4,* http://econ.pstc.brown.edu/faculty/henderson/Wi llAndStuart.pdf or http://www.core.ucl.ac.be/ staff/thisseHandbook/Handbook.html (Accessed 6 June 2004).

Rosenthal, S. S. and Strange, W. C. (2003) Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson and J. F. Thisse (Eds.), *(draft for) Handbook of Urban And Regional Economics, Volume 4*, http://econ.pstc.brown.edu/faculty/henderson/WillAndStuart.pdf.

Statistics New Zealand (2004) Information about the Business Demographic Statistics. *online documentation*, Wellington: Statistics New Zealand [http://www.stats.govt.nz/domino/external/omni/omni.nsf/outputs/Business+Demographic+Statistics; accessed 7 June 2004].