Working Paper Series

02-2013 November 2013

Cluster Techniques for the Classification of Australian Local Governments

Joseph Drew*
** UNE Business School, UNE*

Abstract: Local Government classification systems take several forms typically consisting of groupings of councils based on factors such as the degree of urbanisation, population size and the like. However, to date none of these classification systems have been tested for homogeneity with respect to environmental constraints or validated against external data. This paper employs the New South Wales local government classification system as a representative case study to test its utility and value as a policy instrument. Both theoretical considerations and validation to financial sustainability ratios suggest that classification according to degree of urbanisation and population size produces unsatisfactory results. The theory of environmental constraints is then applied in an exhaustive set of cluster analyses which produce superior local government classifications when validated against financial sustainability ratios.

Keywords: Local government; classification; cluster analysis; Australia

1. Introduction

Many countries possess classification systems for local government which are used as the basis for regulation and policy making. For instance, British policy makers have adopted the Department of Energy, Food and Rural Affairs (DEFRA) sixfold classification system for grouping the different Local Authority Districts and Unitary Authorities within English local government. In an analogous fashion, local government in Australia has official classification systems operating at both the federal and state levels. For example, while the Commonwealth Government employs a 22 group classification system, by contrast the New South Wales (NSW) Government places local authorities into an 11 category system.

Despite the fact that local government classification systems play a pivotal role in local government policy formulation and implementation, with policy instruments typically applied differently to different groups of municipalities, it is surprising that these classification systems have not been tested rigorously for utility and homogeneity. With respect to utility, it has long been known that the best test for a classification system is to compare its performance against variables not used in establishing the various clusters (Aldenderfer and Blashfield, 1984). For instance, in the context of NSW local government, one possibility would be to test the correlation between classification systems and Financial Sustainability Ratios (FSRs).

Existing classification systems almost invariably employ an initial partition of local government entities into different categories based on decisions regarding whether a local authority is urban or rural. Within each of these strata further categories based on population size (and in some instances population density) are then made. Final classification is then frequently based on population size. This paper seeks to add to the existing empirical literature on local government policy making by demonstrating that if a council classification system incorporates more known exogenous constraints, then it can produce more homogenous cluster groups. This in turn may yield classification systems better aligned to the real characteristics of local government systems and thereby provide superior platforms for informed policy making. In so doing, we demonstrate that cluster analysis represents a sound method towards this end.

In essence, using the NSW local government classification system as a case study, this paper shows that cluster analysis can be applied to the classification of local authorities to incorporate exogenous constraints. We use publicly available data on NSW local governments to test the degree of correlation between existing NSW and Australian local government classification systems and FSR developed by the NSW Treasury Corporation (TCorp). Extensive cluster analyses are then conducted utilising a suite of environmental constraint variables, employing the eight most common cluster methods and employing a range of classification group sizes and scales. The resulting clusters are then assessed for their association with FSR. Finally, some lessons and applications of this result are enumerated.

2. Testing Existing NSW and Australian Local Government Classification Systems

Two concurrent systems exist to classify local government in Australia: a national system which classifies local councils into 22 separate groups and seven different state and territory systems, including the NSW Division of Local Government (DLG) system which assigns municipalities into one of eleven discrete groups. Table 1 details the relationship between the Australian and NSW classification schemes. Both methods for classifying local councils employ an initial partition of councils based on judgements regarding the degree of urbanisation, and then subsequent divisions are dominated by population size.

Table 1. NSW and Australian Local Government Classification Systems

Step 1	Step 2	Step 3	Australian code	NSW code
Urban				
	Capital City	Not Applicable	1	1
Pop > 20,000	**Metropolitan Developed**	Small up to 30,000	2	2
Or density > 30/sq. km		Medium 30,000 – 70,000	3	2
Or > 90% LGA pop is urban		Large 70,000 – 120,000	4	3
		Very large > 120,000	5	3
	Regional Town	Small up to 30,000	6	4
		Medium 30,001 – 70,000	7	4
		Large 70,001 – 120,000	8	5
		Very Large > 120,001	9	5
	Fringe	Small up to 30,000	10	6
		Medium 30,001 – 70,000	11	6
		Large 70,001 – 120,000	12	7
		Very Large > 120,001	13	7
Rural				
	Significant Growth	N/A	14	N/a
	Agricultural	Small up to 2,000	15	8
		Medium 2,001 – 5,000	16	9
		Large 5,001 – 10,000	17	10
		Very large 10,001 to 20,000	18	11
	Remote	Extra Small up to 400	19	N/a
		Small 401 – 1,000	20	N/a
		Medium 1,001 – 3,000	21	9
		Large 3,001 to 20,000	22	10

Source: NSW DLG (2012)

However, classification systems of this kind present a number of problems. Firstly, the initial division of urban/rural entities requires subjective judgements (although some guidance is given by population and density thresholds in the NSW context, the 'or 90% of LGA population is urban' clause allows for designation of urban status on a purely subjective basis) which are made all the more difficult by the dynamics of local government urbanization. Other subjective judgements come into play down the classification hierarchy tree. For example, it is not at all obvious that 'regional town', 'fringe', 'agricultural' and 'remote' can be definitively described.

Secondly, many NSW councils occupy very large areas of land that may exhibit traits from a number of qualitative descriptors. For instance, the Tamworth Regional Council in northern NSW, covering 9,893 square kilometres, includes the town of Tamworth which is dominated by retail, residential housing and light industry. However, the majority of its spatial area comprises small settlements, such as Bendemeer and Watsons Creek, which are predominantly service centres for surrounding agricultural activity. This raises the obvious question: should the Tamworth Regional Council be classified as 'regional town' or 'agricultural'? Finally, apparently arbitrary population cut-offs can result in two qualitatively similar councils being assigned into two different classifications. For instance, in NSW the 70,000 population cut-off places Port Stephens Council (population 67,825) in NSW class 4, whereas Coffs Harbour Council (population 72,827) falls in NSW class 5. Moreover, as Port Stephen's population grows, it will move into NSW class 5, without any meaningful change in its other qualitative characteristics (population growth of just over 3% will see it shift up a group).

All of this suggests that the NSW and Australian classification systems may not produce sufficiently homogenous groups of councils suitable for public policy purposes, such as determining the quantum of intergovernmental grants. Furthermore, both systems do not contain any external constraint variables (see section 3), other than population size. It is thus possible that councils within a given classification category face entirely different fiscal and other circumstances.

In this paper we argue that the best test of a classification system is to examine how well it explains variables which are not part of the partitioning scheme (in this case variables other than qualitative judgements and population size) (Aldenderfer and Blashfield, 1984; Zafra-Gomez *et al.*, 2009a). This raises the question of which variables should be employed in a test of this kind. In the present context, it is argued that the NSW TCorp FSR provide a suitable (since the FSR are external to partitioning criteria) and a meaningful (because they have been used to justify proposed large-scale council amalgamations in NSW (ILGRP, 2013)) source of external data for classification system assessment.

FSRs were produced by TCorp in 2013 on the request of the Independent Local Government Review Panel (ILGRP), which was charged with appraising NSW local government performance and proposing remedial policies. The ten ratios produced by TCorp fall into two broad categories; financial performance measures (Operating, Own Source, Cash Expense, Unrestricted Current, Debt Service and Interest Cover ratios) and asset maintenance and renewal measures (Infrastructure Backlog, Asset Maintenance, Asset Renewal and Capital Expenditure ratios). Table 2 lists definitions and measures of central tendency for each of these ratios.

Table 2. Definitions and Measures of Central Tendency of Regression Variables (n = 149)

Variable	Definition	Mean	Standard Deviation
Operating ratio (17.5%)	(operating revenue* - operating expenses) / operating revenue*. Benchmark > -4.0%.	-3.88	9.61
Own Source Revenue ratio (17.5%)	rates, utilities and charges / total operating revenue**. Benchmark > 60.0%.	56.77	13.66
Unrestricted Current ratio (10.0%)	current assets less restrictions / current liabilities less specific purpose liabilities. Benchmark > 1.5.	3.60	2.20
Interest Cover ratio (2.5%)	EBITDA / interest expense. Benchmark > 4.0.	84.13	335.35
Infrastructure Backlog ratio (10.0%)	estimated cost to bring assets to a satisfactory condition / total infrastructure assets. Benchmark < 0.02.	0.13	0.14
Debt Service Cover ratio (7.5%)	EBITDA / (principal repayments + borrowing costs). Benchmark > 2.0.	35.77	270.83
Capital Expenditure ratio (10.0%)	annual capital expenditure / annual depreciation. Benchmark > 1.1.	1.20	0.70
Cash Expense ratio (10.0%)	(current cash and equivalents / (total expenses - depreciation - interest costs)) x 12. Benchmark > 3.0 months.	5.13	4.74
Asset Renewal ratio (7.5%)	Asset renewals / depreciation of building and infrastructure assets. Benchmark > 1.0.	0.66	0.50
Asset Maintenance ratio (7.5%)	actual asset maintenance / required asset maintenance. Benchmark > 1.0.	0.88	0.56
Population	resident population	47,436	58,815
Population density	(individuals/km^2)	757.30	1571.82
% population ATSI	proportion of Aboriginal and Torres Strait Islander population	5.33	7.41
Length of sealed roads	total length of sealed roads maintained by the council (kms)	430.54	254.09
Unsealed roads	total length of unsealed roads maintained by councils (kms)	530.72	580.02

* revenue excludes capital grants and contributions

** revenue includes capital grants and contributions

To test the statistical association between the two existing classification schemes and the FSRs, Pearson correlation coefficients were produced (see Table 3). This linear test is consistent with the functional form for FSRs and population parameter relationships determined by Drew and Dollery (2013). As can be seen, the relationship between classes of councils and FSR is very weak for the majority of ratios (except perhaps the Unrestricted Current and Own Source ratios). However, the high correlation between council group and the Own Source ratio is not surprising given that population size is a major algorithm item in the allocation of federal grant funds - a most significant source of income in NSW local government (NSW Local Government Grants Commission, 2012). Indeed Drew and Dollery (2013) have argued against the inclusion of the Own Source ratio in the FSR reports on these grounds.

Table 3. Correlation Between Classification Systems and FSR

FSR	NSW Classification	Australian Classification
Operating	0.0445	0.0183
Interest Cover	-0.0113	-0.0069
Debt	-0.0755	-0.0678
Unrestricted	0.3431	0.3236
Own Source	-0.6162	-0.6516
Cash Expense	0.1082	0.0907
Infrastructure Backlog	0.0753	0.0706
Asset Maintenance	0.1274	0.1273
Asset Renewal	-0.0579	-0.0575
Capital Expenditure	-0.1212	-0.1212

Poor correlation between council classes and FSRs might have several implications. Firstly, it might suggest that there is little relationship between council characteristics and financial performance (perhaps as a result of poorly conceived or measured FSRs). If this were the case, then we should not expect better correlations in alternate classification schemes. Secondly, there is the possibility that the two variables are indirectly related or have a different functional relationship. This explanation would be inconsistent with existing scholarly analysis (Drew and Dollery, 2013) and should mean that alternate classification schemes will fare little better. Finally, particularly for the case of asset maintenance and renewal ratios, it is possible that alternative accounting approaches may be obscuring associations between council groups and FSRs. In this instance, we may find better associations for ratios with little accrual accounting discretion, but no improvement where ratios involve significant accounting judgements.

A comparison between existing classification systems suggests that the NSW classification method generates marginally closer associations. Moreover, the NSW system only has 11 groups, which is half of that generated by the Australian Government system. This is important given that there is an increasing likelihood of closer correlations as group size increases. To demonstrate this point, we can think about the perfect correlation which would result if 152 unitary groups were constructed. However, the usefulness of classification systems is inversely proportional to group size and policy makers should thus be cognisant of this in choosing between competing classification methods.

3. Cluster Analysis as a Method for Classifying NSW Local Governments

Cluster analysis is a term used to describe a range of empirical techniques specifically designed to classify data into homogenous groups. Its origins lie with the natural sciences and epidemiology. However, it was quite popular 30 or more years ago in the public policy disciplines as a way of stratifying data prior to regression analysis in order to address limited computing power. Whilst other methods of analysis could be adapted to the problem of classifying local governments, such as factor analysis, cluster analysis remains the most appropriate technique for this task. We thus employ it in this paper.

Although cluster analysis has largely lost favour in the public policy disciplines in the face of expanded computing power, contemporary applications still arise. For example, cluster analysis employed in examining the benchmarking value of financial performance data in Spanish local government (Zafra-Gomez *et al*., 2009b). However, in this paper the cluster analysis performed was restricted to just one iterative method (kmeans), it was initially partitioned according to population size (without any supporting evidence to suggest that population size was in fact a relevant determinant), it did not validate the cluster arrangement against external data and it failed to adequately justify the number of clusters assigned (see section 4 for an explanation of kmeans iteration).

The present paper is designed to rectify these weaknesses in an effort to provide a robust method for classifying local governments for applications to benchmarking of FSR (as in the study by Zafra-Gomez et al., 2009b), developing a set of regulatory interventions (see Wilson (2010) for a good account of the UK Comprehensive Performance Assessment program), generating disability factors for a horizontal equalisation grant scheme (Oates, 1999), and the comparative analysis of local government (see NSW Division of Local Government (2012)).

The notion of similarity lies at the core of cluster analysis, whether measured by Euclidian distance (the most common practice and the proximity measure adopted in this paper), angular separation, or some other rule. However, in order to measure similarity between any two councils one must first decide which characteristics will be measured. An analyst cannot simply include every available piece of descriptive data due to the threat of masking variables (see Everitt *et al*., 2001). Thus decisions in this regard must be guided by theoretical insights or by empirical analysis. Applicable theory in this case centres on the concept of environmental constraint. Put simply, environmental constraint is a measure of the external challenges faced by individual councils in providing services to residents. Andrews *et al*. (2005) proposed two rules for selection of environmental constraint variables:

i. They must represent 'degree of difficulty'; and
ii. They must fall beyond the control of local government in the short run (although they may be endogenous in the long run).

The empirical literature on production functions then serves as a guide to the relevant and obtainable variables for consideration. In this regard, Byrnes and Dollery (2002), Drew *et al*. (2012) and Drew and Dollery (2013) suggest (in the context off Australian local government) the use of population size, population density, Aboriginal and Torres Strait Islander (ATSI) demographic data and the length of unsealed and sealed council roads. These variables exhibited the highest modes of statistical significance in recent studies of NSW local government.

Population size has long been considered the best available proxy for approximating local government output and it is ubiquitously employed in empirical work on local government (Drew *et al*., 2013). Population density has also been shown to be an important regressor in econometric models (Ladd, 1992; Holcombe and Williams, 2009) and it is consistent with the premise that the cost of many services (such as garbage collection or provision of sewerage) is dependent, in large part, on distance travelled. ATSI demographic data is a consistent statistically significant regressor in Australian local government cost functions and it is in agreement with international evidence suggesting differing demand for services from certain different population groups (see, for instance, Van De Walle and Van Ryzin, 2011; Drew *et al.*, 2012). Finally, length of council maintained roads is the single largest item of Australian local government expenditure which reflects the fact that local roads represent 80% of Australian road infrastructure (Pricewaterhouse Coopers, 2006). Moreover, different cost functions are associated with sealed (bitumen roads) and unsealed (predominately graded dirt) surfaces (Chakrabarti *et al.*, 2002) thus requiring the inclusion of two differentiated variables. However, it should be stressed that this suite of environmental constraints is thus tailored to meet circumstances in our NSW case study. Accordingly, empirical work in other national contexts should be guided by relevant and recent empirical work in their local government systems.

A final consideration centres on omitted variables. Average wage and unemployment data appear in most fully specified models on the premise that some council services may be normal goods (which holds for many local government systems). However, these variables rarely record statistical significance in the Australian context, not surprising given the focus on 'services to property' in Australian local government compared with the broader 'services to people' content of local government service provision in Europe and North America. Likewise, demographic data on individuals over 65 and under 15, whilst reflective of some specific Australian council services (such as public libraries and local playgrounds) are not major determinants of Australian local government cost structures. Finally, population growth has not been included due to some recent concerns regarding the reliability of intercensal data (see ABS, 2012). Accordingly, these five variables have not been included in the necessarily restricted suite of environmental constraints employed in this study.

Section 4 now sets out the most common clustering methods employed in the scholarly literature, along with an explanation of the rules used for determining the number of groups, initial partitions (for iterative methods only) and validating group composition.

4. Empirical Strategy

Eight of the nine cluster methods commonly included in most statistical software packages are used in this paper (the ninth method - single linkage - is omitted due to its well documented chaining behaviour). Cluster methods can be classified as either hierarchical or iterative. Hierarchical cluster analysis combines single councils into larger and larger homogenous groups as one travels up the hierarchical tree (generally displayed as a dendrogram similar in structure to family trees employed by genealogists), which is a representation of how the group composition changes as the number of groups decreases. These hierarchical methods dominate the cluster analysis literature and include; complete linkage, average linkage, centroid, median linkage, weighted average and Ward's method.

The alternative class of cluster techniques are represented by the iterative methods which start with an initial partition and re-assign individual councils based generally on a measure of central tendency. The major methods in this class are kmeans and kmedians. All eight methods were employed in this analysis (using various group sizes) in order to compare and contrast the application of cluster methods to the problem of local government classification. However, it was apparent from a consideration of the cluster algorithms, set out below, that certain methods (such as the weighted hierarchical techniques) were likely to produce more robust results.

It is difficult to summarising all cluster methods given space constraints. Accordingly, we limit discussion to a consideration of how the Euclidian distance between a single candidate council and an existing group (made up of more than one council) is conceived by the various methods (see, for instance, Aldenderfer and Blashfield, 1984; Everitt *et al.,* 2001). The distance method determines the distance matrix which is instrumental in forming homogenous clusters. In complete linkage, distance is taken as being from the candidate council to the furthest neighbour in the existing group – this approach thus tends to form compact clusters with equal diameters. In contrast, for average linkage, an average is taken of the distance between the candidate council and all councils in the existing group. Average linkage tends to produce clusters with tight variances, thus producing dendrograms with a clear number of clusters. Centroid clustering, on the other hand, does not use a proximity matrix but rather merges councils with the most similar mean vector to an existing cluster (based on a data matrix). It is important to note that both centroid and median linkage provide for the possibility of reversals, thereby preventing the production of a meaningful dendrogram (which, as we see, is a useful tool for determining the number of groups). Weighted average and median linkage are weighted versions of average linkage and centroid, respectively, designed to counter the effect of groups with a larger number of councils being given greater influence. These 'weighted' methods are thus likely to produce better results where there is a significant difference in the number of councils constituting each group. Finally, Ward's method joins a candidate council if it results in the minimum increase to the error sum of squares, which is defined as:

$$ESS = x_i^2 - 1/n \left(\sum x_i\right)^2$$

Due to the minimisation technique employed, Ward's method has been shown to produce poor results if a number of outliers are present.

Iterative methods start with councils partitioned into a pre-determined number of groups. This may be on the basis of a random seed, the results of a known cluster or theoretical consideration. Centroids are then computed for each group and councils are allocated to the group that has the nearest centroid. After a complete pass of the data has been made, new centroids are computed and the iterative process continues until no councils change groups. The difference between kmeans and kmedians is that the latter measures the centroid by the group median (which should be more resistant to skewing), whilst the former uses the arithmetic mean.

A significant problem with these iterative measures is how to determine the number of cluster groups. In this regard (Milligan, 1981) has shown that using the group number indicated by average linkage (which, as noted earlier, produces dendrograms with obvious numbers of clusters) produces robust results. Iterative methods are also sensitive to initial partition and have been observed to produce inconsistent results from random seeding. Accordingly, initial partitions based on hierarchical groupings were also used. Finally, the fact that most software packages do not recompute centres at each assignment of a council to a group (but instead recalculate centroids after an entire pass in order to minimize required computing power) can produce 'orphan group centres', further reducing the efficacy of these iterative methods.

Although the optimal number of classification groups is critical to iterative methods, it also presents a problem when using hierarchical approaches. Ultimately, in hierarchical cluster analysis one is faced with the problem of where to cut the dendrogram tree. We have followed the practice of cutting hierarchical trees at the point prior to a large jump in fusion coefficients (which measure similarity). Whilst more formal tests embodying the fusion jump concept exist, there is a 'lack of consensus about which rule to apply' (Everitt *et al.*, 2001, p77) as well as little consensus on why different results are achieved from different stopping rules on the same data. Accordingly changes of fusion levels, coupled with validation to external data (which we do by assessing the correlation to FSRs), are the most reliable 'stopping' method to employ (Andenderfer and Blashfield 1984).

Scaling is important. If distance matrices were constructed on the raw data, then environmental constraints related to road length and population would overwhelm the influence of measures of the 'degree of difficulty', such as density and ATSI. The method most commonly advocated in the literature involves scaling cluster variables according to range, although z scoring and scaling to the mean are also employed in some studies (see, for instance, Milligan and Cooper, 1988; Granandesikan *et al.,* 1995, Moisl, 2010). Owing to the uncertainty which persists on scaling, we tested all three methods (see Moisl, 2010 for arguments against z scoring and advocating mean scaling; Everitt *et al.* (2001) for caveats against scaling of centroid methods; and Kettenring (2006) for a synoptic discussion of the 'scaling problem'). Our results were consistent with Granandesikan *et al.* (1995) in that validating Pearson FSR correlations were universally stronger with range scaling than z scoring or mean scaling.

Cluster analysis has been criticised with respect to the outcomes achieved in some validation studies as well as its inability to definitively articulate matters such as stop rules (problems which have persisted since Aldenderfer and Blashfield, 1984). Often criticism relates to the poor selection of cluster values, and neglects the need for experimentation with a number of models and the requirement to attempt to validate results (Kettenring, 2006). However, the scholarly literature is also replete with criticisms of multiple regression analysis and its inexpert application (see, for instance, Kennedy, 2003), yet multiple regression analysis remains a dominant technique in applied analysis. The point is that any analytical technique must be used carefully, variables selected according to underlying theoretical foundations and results validated wherever possible. Cluster analysis is no exception.

Validation is undoubtedly a most critical phase of cluster analysis (Kettenring, 2006). One way to gain confidence in a given result is to attempt to replicate the structure using alternate methods. In this regard, cross tabulation of the cluster structures of different methods can provide a certain degree of assurance based on probabilistic considerations. However, as Aldenderfer and Blashfield (1984) have astutely noted, the real test of a classification scheme resides in its ability to explain a set of data not used in the cluster analysis. This test derives from the purpose of classification which is to compartmentalise observations for assimilation, prediction and comparison. Any scheme which fails to have a reasonable level of association to related observations must be considered fallacious.

Accordingly, section 5 details Pearson correlation coefficients for the five most promising classification schemes generated in the analysis. In addition, the cluster structures are cross-tabulated against the existing NSW and Australian classification schemes, which allows for both evaluation of the degree of replication and assessment of the appropriateness of existing partition variables.

5. Outcomes of Cluster Analysis

Table 4 contains the Pearson correlation coefficients for the five most promising cluster structures. The highest improvement in correlation to FSR occurred for 9 group median linkage (6 improvements) followed by 9 group kmedian (5 improvements), 7 group median linkage (4 improvements) and 9 group Ward's and weighted average linkage (both 4 improvements). It should be noted that, for the reasons outlined in section 2, the universal failure of the cluster methods to improve on the Own Source association is of little consequence. In this regard, one might thus consider the number of potential associations to be set at nine. All of these results were recorded after range scaling (autoscaling and mean scaling did not produce high Pearson correlation coefficients). Results were largely predictable given that both existing classifications and predicted clusters contained unequalled numbers of councils. This suggested from the outset that weighted methods would prove superior and that Ward's method might struggle in the presence of some outliers. With respect to kmedians, it is important to note that the successful iteration was based on median linkage initial partitions: the random seed partitions failed to produce useful clusters. This is an important aspect of the use of iterative cluster methods: only when useful initial partitions and group sizes are known can one get replicable and robust results.

Table 4. Correlation Between Cluster Methods and FSR

FSR	Median Linkage (9 groups)	Kmedian (9 groups)	Median Linkage (7 groups)	Ward's (9 groups)	Weighted Average Linkage (9 groups)
Operating	0.1490	0.1013	0.1316	0.1269	0.1160
Interest Cover	0.1094	0.1003	0.1333	0.1209	0.1028
Debt	0.2092	0.1490	0.2267	0.1897	0.1773
Unrestricted	-0.2352	-0.3167	-0.2081	-0.2642	-0.2986
Own Source	0.3790	0.3270	0.3161	0.4827	0.4815
Cash Expense	-0.2600	-0.1548	-0.2307	-0.1979	-0.1939
Infrastructure Backlog	-0.1133	-0.0306	-0.1080	-0.1215	-0.1048
Asset Maintenance	-0.0607	-0.1118	-0.0491	-0.1584	-0.1291
Asset Renewal	-0.0433	-0.0041	-0.0525	-0.0359	-0.0368
Capital Expenditure	0.2141	0.1262	0.1865	0.1449	0.1424

In section 2 we noted a couple of possibilities regarding the poor association between the NSW and Australian classifications and FSRs (either a problem with FSR conception or measurement; a different functional relationship; or the effect of divergent accounting approaches). The significantly higher associations between FSRs and the various classifications proposed by cluster analysis suggests that any measurement issues with FSR must be minimal and that the linear functional form is likely to be valid (consistent with Drew and Dollery, 2013). With respect to the third possibility, continued poor associations with the Infrastructure Backlog and Asset Maintenance ratios implies that differing approaches to the matter of accrual accounting discretion may indeed be obscuring underlying relationships (which is compatible with Woodward, 2007 and TCorp, 2013, p. 48).

A final point related to Table 4 is the generally low variance in correlation coefficients and the consistent sign for the association between cluster groups and FSRs. This is an indication that the results are broadly replicated by the varying cluster analyses reported. It thus provides additional probabilistic evidence in favour of the claim that cluster methods based on environmental constraint factors may provide a useful local government classification scheme.

What this means in practice is that regulatory authorities should dispense with subjective categorisation in favour of empirical approaches which yield homogenous groups of councils facing similar ‘degrees of difficulty’. After all, it is entirely reasonable for local authorities to be benchmarked and compared to other local councils that have similar population size, density, ethnography and road infrastructure challenges irrespective of whether they exist in a nominally ‘rural’ or ‘urban’ setting.

6. Conclusions

The empirical evidence presented in this paper casts considerable doubt over whether existing Australian and NSW state classification schemes do result in homogenous groups of councils which face similar challenges in provision of municipal services. This then suggests that an alternative system may be required. Accordingly, we conducted an exhaustive set of 24 cluster analyses to determine whether a more useful classification scheme could be devised.

Two problems presented themselves in relation to our endeavour. Firstly, the issue of which cluster method should be used, and, secondly the question of which scaling method should be adopted. With regards to the first question we tested eight of the nine principal methods and found that the weighted methods (median linkage and weighted average) produced the most robust results. Ward's method predictably experienced difficulty with the remote council outliers' representative of Australian local government.

In order to address the second problem we conducted cluster analyses using the three methods commonly adopted in the empirical literature – range scaling, autoscaling and scaling to the mean. Our results suggest that range scaling produces the most robust cluster results in the NSW local government application.

In conclusion, this paper presents a valuable insight into the methodological issues associated with applying cluster analysis to local government applications.

References

Australian Bureau of Statistics (ABS). (2012), *Regional Population Growth, Australia 2011-12*, Canberra, ABS.

Aldenderfer, M., and Blashfield, R. (1984), *Cluster Analysis,* London, Sage Publications.

Andrews, R., Boyne, G., Law, J., and Walker, R. (2005), 'External Constraints on Local Service Standards: The Case of Comprehensive Performance Assessment in English Local Government.' *Public Administration*, 83(3), 639-656.

Byrnes, J., and B. Dollery., B. (2002), 'Do Economies of Scale Exist in Australian Local Government? A Review of the Research Evidence1.' *Urban Policy and Research*, 20 (4), 391-414.

Chakrabarti, S., Kodikara, J. and Pardo, L. (2002), 'Survey Results on Stabilisation Methods and Performance of Local Government Roads in Australia.' *Road & Transport Research*, 11(3), 3-16.

Drew, J., and Dollery, B. (2013), 'Is The Game Worth The Candle? Estimating The Impact of the Proposed Greater Sydney Metropolitan Amalgamations on Municipal Financial Sustainability.' *Public Money & Management*, In Print.

Drew, J., Kortt, M., and Dollery, B. (2012), 'Economies of Scale and Local Government Expenditure: Evidence from Australia.' *Administration & Society*. DOI 10.1177/0095399712469191%20.

Drew, J., M. Kortt and Dollery, B. (2013), 'A Cautionary Tale: Council Amalgamations in Tasmania and the Deloitte Access Economics Report.' *Australian Journal of Public Administration*, 72(1), 55-65.

Everitt, B., Landau, S., and Leese, M., (2001), *Cluster Analysis*. London, Arnold.

Granandesikan, R., Kettenring, J. and Tsao, S. (1995), 'Weighting and Selection of Variables.' *Journal of Classification*, 12, 113-136.

Holcombe, R. and D. Williams., D., (2009), 'Are There Economies of Scale in Municipal Government Expenditures?' *Public Finance and Management*. 9 (3):, 416-438.

Independent Local Government Review Panel (ILGRP) (2013), *Future Directions for NSW Local Government - Twenty Essential Steps*, Sydney, ILGRP.

Kennedy, P. (2003), *A Guide to Econometrics*. MIT Press, Cambridge.

Kettenring, J. (2006), 'The Practice of Cluster Analysis.' *Journal of Classification*, 23, 3-30.

Ladd, H. (1992), 'Population Growth, Density and the Costs of Providing Public Services.' Urban Studies, 29(2), 273-295.

Milligan, G. (1981), ‘An Examination of the Effect of Six Types of Error Perturbation of Fifteen Clustering Algorithms.’ Psychometrika, 45, 325-342.

Milligan, G., and Cooper, M. (1988), ‘A Study of the Standardisation of Variables in Cluster Analysis.’ *Journal of Classification*, 5, 181-204.

Moisl, H. (2010), ‘Variable Scaling in Cluster Analysis of Linguistic Data.’ *Corpus Linguistics and Linguistic Theory*, 6(1), 75-103.

NSW Division of Local Government Department of Premier and Cabinet (NSWDLG). (2012), *Comparative Information on NSW Local Government Councils 2010/11*. Sydney, NSWDLG.

New South Wales Local Government Grants Commission (NSWLGGC). (2012), *NSW Local Government Grants Commission 2011-12 Annual Report*. Sydney, NSWLGGC.

Oates, W. (1999), ‘An Essay on Fiscal Federalism.’ *Journal of Economic Literature*, 37(3), 1120.

Pricewaterhouse Coopers (PWC) (2006), *National Financial Sustainability Study of Local Government*. Sydney, Pricewaterhouse Coopers.

TCorp (2013), *Financial Sustainability of the New South Wales Local Government Sector*, Sydney, TCorp.

Van De Walle, S. and Van Ryzin, G. (2011), ‘The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons From a Split-Ballot Experiment.’ *Public Administration*, 89(4), 1436-1450.

Wilson, J. (2010), ‘Comprehensive Performance Assessment – Springboard or Deadweight?’ *Public Money & Management*, 24(1), 63-68.

Woodward, R. (2007), ‘Discussion of Jones and Walker.’ *ABACUS*, 43(3), 419-427.

Zafra-Gomez, J., Lopez-Hernandez, A., and Hernandez-Bastida, A. (2009a), ‘Developing and Alert System for Local Governments in Financial Crisis.’, Public Money & Management, 29(3), 175-181.

Zafra-Gomez, J., Lopez-Hernandez, A., and Hernandez-Bastida, A. (2009b), ‘Evaluating Financial Performance in Local Government: Maximising the Benchmarking Value.’, *International Review of Administrative Sciences*, 75(1), 151-167.