Journal of Modern Applied Statistical Methods

Volume 11 | Issue 2 Article 18

11-1-2012

Small-to-Medium Enterprises and Economic Growth: A Comparative Study of Clustering Techniques

Karim K. Mardaneh
University of Ballarat, Mount Helen, Australia

Journal of Modern Applied Statistical Methods
November 2012, Vol. 11, No. 2, 469-478

1538 – 9472/12/$95.00

Small-to-Medium Enterprises and Economic Growth: A Comparative Study of Clustering Techniques

Karim K. Mardaneh
University of Ballarat,
Mount Helen, Australia

Small-to-medium enterprises (SMEs) in regional (non-metropolitan) areas are considered when economic planning may require large data sets and sophisticated clustering techniques. The economic growth of regional areas was investigated using four clustering algorithms. Empirical analysis demonstrated that the modified global *k*-means algorithm outperformed other algorithms.

Key words: Clustering, *k*-means, Ward's clustering, firm size, industry structure, regional economy.

Introduction

Clustering techniques are compared by examining the relationship between industry structure and business size with economic growth using Australian regional areas (non-metropolitan) data. Pagano (2003) examined firm size and industry structure; however, the study did not consider in combination the role of both industry structure and size of business in economic growth. This study uses four clustering techniques on statistical local area (SLA) regions to examine the performance of these clustering methods on small-to-medium enterprises (SMEs) data sets. Researchers such as Beer and Clower (2009) have used clustering techniques for pattern recognition; however, there is a gap in the literature in terms of applying clustering methods to SMEs related problems.

Data mining facilitates the identification of useful information within data reservoirs and involves the application of discovery algorithms to the data. Cluster analysis is an important data mining task (Mardaneh, 2007). Cluster analysis has been used by contemporary researchers when the number of observations in a particular field is fairly large (Freestone, Murphy and Jenner, 2003). This study adopts cluster analysis and uses four methods of clustering algorithms: Ward's (Ward, 1963), the *k*-means (Hartigan & Wong, 1979), global *k*-means (Likas, Vlassis & Verbeek, 2003), and the modified global *k*-means (Bagirov, 2008; Bagirov & Mardaneh, 2006). These algorithms are employed to cluster SLAs based on industry structure and size of the businesses within those areas and to compare the function of the algorithms to identify a clustering algorithm that is most suitable for clustering SMEs data. This study addresses the gap in understanding the combined role of industry structure and size of business in economic growth, as well as the cluster analysis of the SMEs data sets.

Karim Ken Marandeh is a Lecturer and Post Doctoral Fellow in the Business School. His research interests include regional economics, economy of small to medium size enterprises, data mining and its application in economics and business problems. Email him at: k.mardaneh@ballarat.edu.au.

Literature Review

Understanding economic growth requires a thorough consideration of the role of industry structure and the size of business (micro, small, medium or large). Regions with an industry structure that enables wealth-creating initiatives will have a better economic condition (Delgado, et al., 2010). In addition, the distribution of a region's economic activity across industries is considered to be a major determinant of the resilience of its economy (Australian Government Department of

Transport and Regional Services, 2003). Previous studies in this area mainly focus on formation and growth (Dobbs & Hamilton, 2007; Mueller, et al., 2008; Hudson, et al., 2001; Beugelsdijk, 2007; Sierdjan, 2007; Koster, 2007; Armington & ACS, 2002; Pagano & Schivardi, 2000, Dejardin & Fritsch, 2010), as well as organizational attitude to change, and success and failure (Walker & Brown, 2004; Agarwal & Audretsch, 2001; Gray, 2002; Feser, et al., 2008; Dejardin & Fritsch, 2010). A few studies have investigated industry structure and size of business (Okamuro, 2006, Okamuro & Hobayashi, 2006; Pagano, 2003; Pagano & Schivardi, 2000); however, these studies did not identify the drivers of economic growth in relation to those factors.

Clustering, or cluster analysis, is a challenging problem for which different algorithms have been proposed. Cluster analysis addresses the problem of organizing a collection of patterns or objects into clusters based on similarity so that objects in the same cluster are in some way more similar to each other than to those in other clusters (Bagirov, 2008; Bagirov & Mardaneh, 2006). Beer and Maude (1995) used cluster analysis to examine changes in economic functions of towns, and Smith (1965) used clustering in the study of economic functions of Australian regional towns. In this study the clustering technique is used to collect SLAs into clusters so that SLA regions within a cluster are similar to each other and are different from regions in the other clusters.

Clustering methods in general have been used in business and economics disciplines. Ward's clustering method has been widely used in consumer behavior (Greeno, Sommers & Kernan, 1973; Kernan & Bruce, 1972), marketing and economics studies (Eliashberg, Lilien & Kim, 1995; Blin & Cohen, 1997; Doyle & Saunders, 1985). Ward's clustering in particular has been used to study Australian regional economic development (Beer & Maude, 1995; Beer & Clower, 2009; Sorensen & Weinand, 1991), urbanization in Australian economy (Freestone, et al., 2003), and marketing themes and strategies (Ho & Hung, 2008; Wong & Saunders, 1993). Unlike the Ward's method, the *k*-means algorithms have not been widely used within these disciplines. The *k*-means algorithms have been mainly used in information technology and data mining studies and in a few marketing studies (Calantone & Sawyer, 1978; Moriarty & Venkatesan, 1978; Schaninger, Lessig & Panton, 1980). The *k*-means algorithm has only recently been used in regional economics studies (Mardaneh, 2012).

This study explores whether the *k*-means algorithm and its variations could provide a better tool for regional economics studies than the Ward's clustering algorithm. The framework of the study is based on SMEs, the two variables (industry structure and business size) and the comparative experiment of the four algorithms. A more efficient algorithm that better clusters the SMEs data could help advance the understanding of industry structure and size of business (SMEs) which, in turn, could provide valuable information regarding regional economic growth.

Methodology

Using regional Australian data this study examines the influence of the industry structure and size of businesses on the economic growth of SLAs. To measure growth, individual weekly income is used as a proxy for economic growth and assumes that SLAs with more people in \$1,000-\$1,999 and \$2,000 and over income per week must enjoy a particular industry structure and business sizes. To investigate this, SLAs based on industry structure and three business sizes (micro, small, medium) are clustered. Clustering is conducted three times, once for each size of business, using the *k*-means, global *k*-means, modified global *k*-means and Ward's clustering algorithms. Results are compared to identify the clustering algorithm that is most suitable for clustering the SMEs data.

Data for this study is obtained from the Australian Bureau of Statistics (ABS, 2007) and uses information from the *Counts of Australian Businesses, including Entries and Exits, June 2003-June 2007*, which includes *Businesses by Industry Division by SLA by Employment size ranges*. This is provided as categories of data for businesses by industry division (see Appendix A for the list of industries). The data exhibits sixteen industry types and the number of employees at each SLA based on business size.

The ABS classifies size of businesses as micro business (1-4 employees), small business (5-19 employees), medium business (20-199 employees) and large business (200 and over employees). This classification is maintained herein, however, this study does not include large businesses (200 and over employees) because the relevant data were too sparse. For the same reason the 'electricity, gas and water supply' industry is excluded from the analysis. Because this study focuses on regional geographical areas - and due to the fact that the industry structure and number of business sizes in regional areas are very different from metropolitan areas - metropolitan data is excluded; this avoids skewness in analysis. After removing metropolitan SLAs and outliers (extreme values in data set) 661 regional SLAs were included in the analysis.

The percentage of people in two weekly income levels, \$1,000-\$1,999 and \$2,000 and over, are considered per SLA. The median of the percentage for each income level is calculated across all SLAs (11.8% and 1.9% for each income level, respectively). SLAs above median within both income levels are considered as SLAs having a higher level of economic growth and are labeled as category 1; the remaining SLAs are considered as SLAs with a lower level of economic growth and are labeled as category 2.

Samples in the data are comprised of SLAs under three business sizes (1-4, 5-19, 20-199) and fifteen industry types which form the data set. To identify the industry type(s) and business size(s) with higher or lower levels of contribution to the economic growth of a SLA (which allocates a SLA to categories 1 or 2) clustering analysis was conducted using three SMEs data sets (see Tables 1-3). For this, the *k*-means, global *k*-means, modified global *k*-means and Ward's clustering algorithms were applied (see Tables 4-6).

Clustering Algorithms

Clustering algorithms can be used to analyze large data sets comprising a myriad of economic and social variables. They seek to group samples with similar characteristics and ensure maximum statistical separation from other contrasting clusters. This is a process of pattern recognition which simplifies understanding of large data sets. In one classification, clustering algorithms are classified as hierarchical or iterative algorithms. Hierarchical methods begin with a set of clusters and place each sample in an individual cluster. Clusters are then successively merged to form a hierarchy of clusters (Guha, et al., 2001). Iterative methods start by dividing observations into some predetermined number of clusters. Observations are then reassigned to clusters until some decision rule terminates the process (Punj & Stewart, 1983). Ward's clustering algorithm is hierarchical, while the *k*-means and its variations are iterative.

Ward's Algorithm

Ward's algorithm seeks to group a set of n members, which are called subsets or groups in relation to an objective function value. The method seeks to unite two of the n subsets to reduce the number of subsets to $n-1$ in a way that minimizes the change in the objective function's value. The $n-1$ resulting subsets are examined to determine if a third member should be grouped with the first pair. If necessary this procedure can be continued until all n members of the original array are in one group (Ward, 1963).

The *k*-means Algorithm

The *k*-means algorithm considers each sample (SLAs in this study) as a point in n-dimensional space (R^n) and chooses k centers (also called centroids) and assigns each point to the cluster nearest the center. The center is the average of all points in the cluster, that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. The *k*-means algorithm is an efficient clustering algorithm, but it is sensitive to the choice of starting points (Bagirov, 2008).

The Global *K*-means Algorithm

The global *k*-means algorithm was proposed to improve global search properties of *k*-means algorithms. The global *k*-means algorithm (Likas, et al., 2003) computes clusters successively. At the first iteration of this algorithm the centroid of a set A is computed and, in order to compute k-partition of the k^{th} iteration, the algorithm uses centers of $k-1$ clusters from the previous iteration (Likas, et al., 2003).

The Modified Global K-means Algorithm

The modified global k means algorithm computes clusters incrementally and, to compute the k-partition of a data set, it uses $k-1$ cluster centers from the previous iteration. An important step in this algorithm is the computation of a starting point for the k^{th} cluster center. This starting point is computed by minimizing the so-called auxiliary cluster function. (Bagirov, 2008; Bagirov & Mardaneh, 2006)

Empirical studies of the performance of clustering algorithms (Punj & Stewart, 1983) suggest that one of the iterative clustering methods (e.g., k-means clustering) is preferable to hierarchical methods (e.g., Ward's clustering). The k-means appears to be more efficient (Mezzich, 1978; Milligan, 1980; Bayne, et al., 1980) if a non-random starting point is specified. When a clustering algorithm includes more and more observations its performance tends to deteriorate: This effect may be the result of outliers entering into the solution. The k-means appears to be more robust than hierarchical methods with respect to the presence of outliers. Results from this study suggest that the more efficient version of the k-means algorithm (modified global k-means) may better cluster SMEs data and could help with further understanding of industry structure and the size of business in regional economics studies.

Results

The analysis clustered SLAs based on industry type and three business sizes (1-4, 5-19, 20-199). Industry, cluster category and the cluster centroids values are reported in Tables 1-3; industries are reported in the tables only if the difference between cluster centroids values in cluster category 1 and 2 for a particular industry is 0.1 or more. In addition, industry type and size of business (variables) with higher cluster centroids value in cluster category 1 are considered as variables with a higher level of contribution to economic growth. Industry type and size of business with higher cluster centroids value in cluster category 2 are considered as variables with a lower level of contribution to economic growth.

As shown in Tables 1-3, the construction, retail trade and personal and other services industries indicate a higher level of contribution to economic growth in all three firm sizes. By contrast, the agriculture, forestry and fishing and wholesale and communication services industries show a lower level of contribution to economic growth in all three firm sizes.

The property and business services industry shows a higher level of contribution to economic growth for both firm sizes 1-4 and 5-19; however, this industry shows a lower level of contribution to economic growth in firms sized 20-199. The cultural and recreational services industry shows a higher level of contribution in both 1-4 and 20-199 sized firms, but shows a lower level of contribution for those sized 5-19. The transport and storage industry shows a higher level of contribution for both 5-19 and 20-199 sized firms; however, it shows a lower level of contribution for 1-4 sized firms.

The health and community services industry shows a higher level of contribution for firms sized 5-19, but shows a lower level of contribution for firms sized 1-4. The finance and insurance industry shows a higher level of contribution only for larger firms, size 20-199. The accommodation, cafes and restaurants industry shows a higher level of contribution for sized 20-199 firms; however, it shows a lower level of contribution for both of the other two sizes. The mining and manufacturing industries both show a lower level of contribution for 1-4 and 20-199 sized firms.

By applying clustering analysis, this study sought to identify the most efficient algorithm for clustering SMEs data. For this, the objective function value and the CPU time spent by each algorithm for clustering were calculated. Clustering was conducted for 2, 5, 10, 15, and 20 cluster numbers for comparison. The analyses in this study were conducted using an Intel Core 2 Duo, 2.99 GHz, PC. Tables 4- 6 show the number of clusters (N), values of the objective function ($f \times 10^5$) and CPU time spent for the

Table 1: Higher/Lower Level of Industry Contribution in Economic Growth; Firm Size 1-4

Industry	Cluster Category	
	1	2
	Cluster Centroids	
Higher Level of Contribution to Economic Growth		
• Construction	17.42	13.62
• Retail Trade	14.33	11.98
• Property and Business Services	12.48	11.55
• Personal and other Services	3.68	3.55
• Cultural and Recreational Services	1.61	1.41
Lower Level of Contribution to Economic Growth		
• Mining	0.51	0.76
• Communication Services	1.54	1.90
• Wholesale	3.81	4.09
• Accommodation, Cafes and Restaurants	3.53	4.50
• Health and Community Services	4.79	4.92
• Manufacturing	4.98	5.13
• Transport and Storage	5.84	6.10
• Agriculture, Forestry and Fishing	21.32	26.81

Table 2: Higher/Lower Level of Industry Contribution in Economic Growth; Firm Size 5-19

Industry	Cluster Category	
	1	2
	Cluster Centroids	
Higher Level of Contribution to Economic Growth		
• Retail Trade	17.64	14.88
• Construction	12.41	8.35
• Property and Business Services	11.43	10.83
• Transport and Storage	5.07	4.87
• Health and Community Services	5.01	4.88
• Personal and other Services	3.57	3.40
Lower Level of Contribution to Economic Growth		
• Communication Services	0.60	0.77
• Cultural and Recreational Services	1.70	1.88
• Wholesale	3.99	5.10
• Accommodation, Cafes and Restaurants	7.50	7.69
• Agriculture, Forestry and Fishing	20.99	27.97

Table 3: Higher/Lower Level of Industry Contribution in Economic Growth; Firm Size 20-199

Industry	Cluster Category	
	1	2
	Cluster Centroids	
Higher Level of Contribution to Economic Growth		
• Retail Trade	15.89	13.69
• Accommodation, Cafes and Restaurants	11.73	11.03
• Construction	10.17	5.5
• Transport and Storage	6.14	4.73
• Cultural and Recreational Services	3.50	3.26
• Finance and Insurance	2.00	1.39
• Personal and other Services	1.50	1.30
Lower Level of Contribution to Economic Growth		
• Communication Services	0.41	0.98
• Mining	0.66	1.19
• Wholesale	4.39	5.72
• Manufacturing	8.80	10.11
• Property and Business Services	10.11	11.35
• Agriculture, Forestry and Fishing	16.91	20.93

analysis (t) for the multi-start k-means (MSKM), global k-means (GKM), modified global k-means (MGKM), and Ward's (WARD) clustering algorithms. Results from the analysis, including the objective function values and CPU time spent for the calculation by each algorithm, are shown in Tables 4-6.

Algorithm Performance

Results presented in Table 4 show that MGKM algorithm outperforms both the MSKM and GKM when the number of clusters $N \geq 10$. Regardless of the number of clusters, the MGKM outperforms WARD and WARD gives the worst results compared to all other algorithms. The GKM requires less CPU time; however, its solutions are not better. MGKM requires more CPU time, particularly when the number of clusters increases ($N \geq 10$). Similarly CPU time for MSKM and GKM increases as the number of clusters N increases. CPU time for WARD is nearly constant for any cluster number N because it is a hierarchical algorithm and, unlike the other three algorithms, it does not go through iterations.

Results in Table 5 show that the MGKM algorithm outperforms both MSKM and GKM when the number of clusters $N \geq 10$. With any number of clusters MGKM outperforms WARD. MGKM requires more CPU time particularly when the number of clusters increases ($N > 5$). Similarly CPU time for MSKM and GKM increases as the number of clusters N increases. CPU time for WARD is almost constant for any cluster number N. Table 6 shows that in some cases, for example, $N = 2, 15$, MSKM performed slightly better than MGKM, however, the difference in performance is minimal. With any number of clusters MGKM outperforms WARD. MGKM required more CPU time for all cluster

Table 4: Data Set 1* - Comparative Values for Algorithms;
Firm Size 1-4

N	MSKM		GKM		MGKM		WARD	
	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t
2	7.582	0.01	7.582	0.01	7.582	0.03	8.274	0.20
5	5.018	0.07	5.018	0.07	5.018	0.12	5.471	0.18
10	3.747	0.12	3.721	0.17	3.721	0.29	4.054	0.18
15	3.111	0.20	3.044	0.26	3.025	0.45	3.268	0.18
20	2.617	0.32	2.542	0.39	2.549	0.57	2.759	0.18

*Data Set 1 includes micro businesses with 1-4 employees across 15 industry types.

Table 5: Data Set 2* - Comparative Values for Algorithms;
Firm Size 5-19

N	MSKM		GKM		MGKM		WARD	
	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t
2	8.721	0.00	8.721	0.01	8.721	0.03	9.122	0.18
5	5.955	0.04	5.944	0.07	5.944	0.12	6.376	0.18
10	4.331	0.10	4.355	0.18	4.341	0.29	4.705	0.18
15	3.609	0.23	3.605	0.28	3.570	0.45	3.888	0.20
20	3.208	0.31	3.201	0.39	3.133	0.62	3.413	0.18

*Data Set 2 includes small businesses with 5-19 employees across 15 industry types.

Table 6: Data Set 3* - Comparative Values for Algorithms;
Firm Size 20-199

N	MSKM		GKM		MGKM		WARD	
	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t	$f\times10^5$	t
2	15.929	0.01	15.930	0.01	15.930	0.03	16.453	0.18
5	11.488	0.03	11.058	0.09	11.058	0.12	12.164	0.17
10	7.811	0.10	7.811	0.18	7.814	0.28	8.818	0.18
15	6.324	0.15	6.336	0.28	6.345	0.43	7.029	0.18
20	5.607	0.32	5.494	0.35	5.513	0.60	6.062	0.18

*Data Set 3 includes medium businesses with 20-199 employees across 15 industry types.

numbers *N*. CPU time for GKM and MGKM increased as the number of clusters *N* increased. CPU time for WARD is nearly constant for any cluster number *N*.

Identifying a clustering algorithm that could help with more efficient cluster analyses of SMEs data is important. A more efficient clustering algorithm may help provide a more accurate and precise grouping of the data points (in this study, geographical areas) based on their similarity. This, in turn, will help with identifying shared characteristics between members (data points) of a cluster. Understanding these characteristics provides a diagnostic of the factors that generate those characteristics. As this study shows, such an understanding can help with identifying the role that each combined industry and business size could play in the economic growth or decline of geographical areas and also whether they have higher or lower contributions to the economic growth of an area.

Conclusion

Cluster analysis revealed clusters of industries associated with industry structure and size of business. This study presented numerical results from three data sets. The results clearly show that the modified global *k*-means algorithm is more efficient for solving clustering problems in SMEs data sets; this algorithm outperforms multi-start *k*-means, global *k*-means and Ward's clustering algorithms. The modified global *k*-means algorithm, however, requires more computational efforts than the global *k*-means algorithm, but is the most promising among all tested algorithms.

The findings from this study provide an improved method for clustering using a more efficient algorithm and, as a result, provide a better understanding of industry structure and size of businesses in regional areas. These findings have policy implications for future economic planning and focus on SMEs for regional areas and will provide paths in identifying significant factors that require further investigation using qualitative methods to ascertain the importance of the clusters and their relationship to SMEs.

References

Agarwal, R., & Audretsch, D. B. (2001). Does Entry Size Matter? The impact of the life cycle and technology on firm survival. *The Journal of Industrial Economics*, *49*, 21-43.

Armington, C., & Acs, Z. J. (2002). The determinants of regional variation in new firm formation. *Regional Studies*, *36*, 33-45.

Australian Bureau of Statistics. (2007). *Counts of Australian businesses, including entries and exits, June 2003 to June 2007*, Cat. No. 8165.0. Canberra: ABS.

Australian Government Department of Transport and Regional Services. (2003). *Information Paper 49: Focus on Regions: No.1: Industry structure*, 1-67.

Bagirov, A. M. (2008). Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, *41*, 3192-3199.

Bagirov, A. M., & Mardaneh, K. (2006). Modified global *k*-means algorithm for clustering in gene expression data sets. In *Proceedings of 2006 Workshop on intelligent systems for bioinformatics (WISB 2006)*, M. Boden. & T. L. Bailey (Eds.), 23-28. Hobart: Australian Computer Society (ACS).

Bayne, C. K., Beauchamp, J. J., Begovich C. L., & Kane, V. E. (1980). Monte Carlo comparisons of selected clustering procedures. *Pattern Recognition*, *12*, 51-62.

Beer, A., & Clower, T. (2009). Specialization and growth: Evidence from Australia's regional cities. *Urban Studies*, *46*, 369-388.

Beer, A., & Clower, T. (2009). Specialization and growth: Evidence from Australia's regional cities. *Urban Studies*, *46*, 369-388.

Beer, A., & Maude, A. (1995). Regional cities in the Australian urban system, 1961-1991. *Urban Policy and Research*, *13*, 135-148.

Beugelsdijk, S. (2007). Entrepreneurial culture, regional innovativeness and economic growth. *Journal of Evolutionary Economics*, *17*, 187-210.

Blin, J. M., & Cohen, C. (1997). Technological similarity and aggregation in input-output systems: A cluster-analytic approach. *The Review of Economics and Statistics*, *59*, 82-91.

Calantone, R. J., & Sawyer, A. G. (1978). The stability of benefit segments. *Journal of Marketing Research*, *15*, 395-404.

Dejardin, M., & Fritsch, M. (2010). Entrepreneurial dynamics and regional growth. *Small Business Economics*, *36*, 377-382.

Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. *Journal of Economic Geography*, *10*, 495-518.

Dobbs, M., & Hamilton, R. T. (2007). Small business growth: recent evidence and new directions. *International Journal of Entrepreneurial Behaviour & Research*, *13*, 296-322.

Doyle, P., & Saunders, J. (1985). Market segmentation and positioning in specialized industrial markets. *Journal of Marketing*, 24-32.

Eliashberg, J., Lilien, G. L., & Kim, N. (1995). Searching for generalization in business marketing negotiations. *Marketing Science*, *14*, 47-60.

Feser, E., Renski, H., & Goldstein, H. (2008). Clusters and economic development outcomes: An analysis of the link between clustering and industry growth. *Economic Development Quarterly*, *22*, 324-344.

Freestone, R., Murphy, P., & Jenner, A. (2003). The functions of Australian towns revisited. *Tijdschrift voor Economische en Sociale Geografie*, *94*, 188-204.

Gray, C. (2002). Entrepreneurship, resistance to change and growth in small firms. *Journal of Small Business and Enterprise Development*, *9*, 61-72.

Greeno, D. W., Sommers, M. S., & Kernan, J. B. (1973). Personality and implicit behavior patterns. *Journal of Marketing Research*, *10*, 63-69.

Guha, S., Rastogi, R., & Shim, K. (2001). Cure: An efficient clustering algorithm for large databases. *Information Systems*, *26*, 35-58.

Hartigan, J. A., & Wong, M. A. (1979). A *k*-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*, 100-108.

Ho, H-F., & Hung, C-C. (2008). Marketing mix formulation for higher education: An integrated analysis employing analytic hierarchy process, cluster analysis and correspondence analysis. *International Journal of Educational Management*, *22*, 328-340.

Hudson, M., Smart, A., & Bourne, M. (2001) Theory and practice in SME performance measurement systems. *International Journal of Operations & Production Management*, *21*, 1096-1115.

Kernan, J. B., & Bruce, G. D. (1972). The socioeconomic structure of an urban area. *Journal of Marketing Research*, *9*, 15-18.

Koster, S. (2007). The entrepreneurial and replication function of new firm formation. *Tijdschrift voor Economische en Sociale Geografie*, *98*, 667-674.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*, 451-461.

Mardaneh, K. (2007). *Optimization based clustering and classification algorithms in analysis of microarray gene expression data sets*. Ph.D. thesis, School of Information Technology and Mathematical Sciences, University of Ballarat.

Mardaneh, K. (2012). A Study of population change via clustering of australian regional areas: An optimisation approach. *Australasian Journal of Regional Studies*, *1*, 257-280.

Mezzich, J. E. (1978). Evaluation clustering methods for psychiatric diagnosis. *Biological Psychiatry*, *13*, 265-281.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325-342.

Moriarty, M., & Venkatesan, M. (1978). Concept evaluation and market segmentation. *Journal of Marketing*, *42*, 82-86.

Mueller, P., Stel, A. V., & Storey, D. J. (2008). The effects of new firm formation on regional development over time: The case of Great Britain. *Small Business Economics*, *30*, 59-71.

Okamuro, H. (2006). Determinants of R&D activities by start-up firms: Evidence from Japan. *Institute of Economic Research, Hitotsubashi University*, Working Papers Series, No. 2006-2.

Okamuro, H., & Hobayashi, N. (2006). The impact of regional factors on the start-up ratio in Japan. *Journal of Small Business Management*, *44*, 310-313.

Pagano, P. (2003). Firm size distribution and growth. *Scandinavian Journal of Economics*, *105*, 255-274.

Pagano, P., & Schivardi, F. (2000). *Firm size distribution and growth, mimeo*. Retrieved December 10, 2010 from: http://digilander.libero.it/fshivardi.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, *10*, 134-148.

Schaninger, C. M., Lessig, V. P., & Panton, D. B. (1980). The complementary use of multivariate procedures to investigate nonlinear and interactive relationship between personality and product usage. *Journal of Marketing Research*, *17*, 119-124.

Sierdjan, K. (2007). The entrepreneurial and replication function of new firm formation. *Tijdschrift voor Economische en Sociale Geografie*, *98*, 667-674.

Smith, R. H. T. (1965). Method and purpose in functional town classification. *Annals of the Association of American Geographers*, *55*, 539-548.

Sorensen, T., & Weinand, H. (1991). Regional well-being in Australia revisited. *Australian Geographical Studies*, *29*, 42-70.

Walker, E., & Brown, A. (2004). What success factors are important to small business owners? *International Small Business Journal*, *22*, 577-594.

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.

Wong, V., & Saunders, J. (1993). Business orientation and corporate success. *Journal of Strategic Marketing*, *1*, 20-40.

Appendix A: List of Industries

Industry Types	
Agriculture, Forestry and Fishing	Transport, and Storage
Mining	Communication Services
Manufacturing;	Finance and Insurance
Electricity, Gas, and Water Supply	Property and Business Services
Construction	Education
Wholesale Trade	Health and Community Services
Retail Trade	Cultural and Recreational Services
Accommodation, Cafes and Restaurants	Personal and Other Services

Source: (ABS, 2007)