A STUDY OF POPULATION CHANGE VIA CLUSTERING OF AUSTRALIAN REGIONAL AREAS: AN OPTIMISATION APPROACH

Karim Mardaneh
The Business School, University of Ballarat, University Drive, Mount Helen, Victoria, 3353, Australia. Email: k.mardaneh@ballarat.edu.au

ABSTRACT: Grouping regional towns and cities in Australia according to economic functions could improve understanding of the importance of economic factors in determining growth. Several researchers have used clustering techniques to examine the growth and characteristics of regional cities in Australia. The current study extends clustering methodologies by adopting an optimisation approach based on a clustering technique using the *k*-means algorithm to investigate the impact of socio-economic factors on population growth and decline in regional Australia. The analysis in the paper suggests that industry of employment, individual weekly income, age group and education level have an important impact on population change. These findings have policy implications for economic planning of regional areas in Australia.

KEY WORDS: Population, Clustering, Optimisation, Regional Australia, *k*-means algorithm

ACKNOWLEDGEMENTS: The author acknowledges the support and advice provided by Associate Professor Jerry Courvisanos, Ms. Alison Ollerenshaw, Dr. Matthew Rockloff, Mr. David Lynch and Dr. Jackie Tuck in the preparation of this paper. The author would also like to thank the anonymous reviewers and the editor of the Australasian Journal of Regional Studies for their recommendations for improving this paper.

1. INTRODUCTION

The current study investigates the impact of several socio-economic factors on population growth and decline in regional Australian Statistical Local Areas (SLAs) within the period 2001-2006. Studies of this type are few in number, but this study builds on previous clustering studies (e.g. Beer and Clower, 2009). Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters (Bagirov, 2008). The so-called *k*-means clustering algorithm (one of the algorithms that can be used in cluster analysis) has been used here. This approach uses multiple regression analysis to determine the impact of the generated optimal clusters on population growth.

Greater government and research attention has been directed towards regional and rural Australia in recent years. Much of this attention has been in response to the population migration from rural and regional areas to urban centres, with over two-thirds of Australia's population now residing in a capital city SLA (Australian Bureau of Statistics (ABS), 2010b). In 2008-09, population growth rates in Australian capital city statistical divisions averaged 2.3%, which exceeded the growth rate of the remainder of Australia (1.9%). Conversely, rural Australia is experiencing population losses, especially in north-eastern and south-eastern Australia (ABS, 2010b). This pattern can be observed in the regional west of Victoria, for instance, where a clear decline in population is most apparent in Local Government Areas located farthest from Melbourne, whose declining populations are rapidly ageing.

Population migration presents significant challenges to Australia's major cities as well as to rural Australia (Infrastructure Australia, 2010). Indeed, regional Australia's economic contribution is essential to the nation's current and future prosperity (Beer and Maude, 1995). Approximately two-thirds of Australian export earnings come from Australian regional industries such as agriculture, mining and manufacturing, as well as tourism and the retail sector, service and manufacturing (Australian Government Department of Regional Australia, Regional Development and Local Government, 2010).

A recent report to the Standing Committee on Regional Development for the Regional Development Council, (2010) suggests that there is a significant challenge in attracting and retaining skilled workers in regional Australia. Development strategies must be tailored to the differing needs and resources of particular areas. Regional Development Australia (2010) reports on an Organisation for Economic Co-operation and Development (OECD) study which discusses in particular the need to encourage regional competitiveness by focusing on the development of underutilised resources. Strategies for doing this include promoting and harnessing regional innovation through tertiary education. Courvisanos (2009) likewise identifies the crucial role of innovation and learning for sustainable development in regional Australia, a crucial factor in limiting decline in some regions and encouraging population growth in stronger regions.

2. LITERATURE REVIEW

This review of the literature relevant to this topic and this paper provides a brief summary based on three broad parameters: population change, economic activity, and analytical methodology.

Population change has been the focus of research for a long time. Many studies have considered the influence of a range of socio-economic factors on population change. A few of these are outlined here. Adelman (1963) investigated the impact of factors such as population density, per capita income, and an educational index on variations in fertility and mortality. Frisbie and Poston (1975, 1978) examined population change against a number of socio-economic factors such as income, employment, racial composition, age, and proximity to metropolitan areas. Shumway and Davis (1996) also used sustenance activities (agriculture, mining and other primary industries) and other factors such as age and household income to examine the impact of each on net migration. Goetz and Debertin (1996) investigated the impact of land price, income and employment on rural population decline in the U.S. Millward (2005) examined the impact of factors such as the unemployment rate, household income, resource-industry employment, proximity to a major urban centre, population density, and commuting on population growth in Nova Scotia, Canada. Using a Canadian case study, Polese and Shearmur (2006) considered resource dependency, distance, income, human resources, and industrial mix in an investigation about regional decline. Findings of these above mentioned studies are in some cases conflicting (e.g. income having positive or negative impact on population change). They indicate that in most cases population growth and decline is impacted directly by factors such as level of income, employment, age, and is inversely impacted by industries such as agriculture, mining, and wholesale. The most common factors identified in these studies are age and income followed by the second group of factors of employment and industry types.

Factors influencing regional *economic activit*y have also received much attention from academic researchers. Research into the economic functions of rural Australia goes back to Smith's (1965) study that used industry of employment data to cluster towns with a similar functional specialisation. In more recent years similar research has been conducted by Beer and colleagues (Beer *et al.*, 2003; Beer and Keane, 2000). Beer and Maude (1995) examined changes in the economic functions of Australian regional cities and the relationship between industry structure and population change rates. They also compared population change and the labour force growth rates with national trends using cluster analysis to examine changes in their economic functions between 1961 and 1991.

O'Connor *et al*. (2001, p. 17) show that "in order to understand the change in contemporary Australian society, it is necessary to understand the structural changes in the economy" brought about by new technologies, new products and the evolution of social attitudes. The sustainability of regional cities and areas also hinges on such factors as transportation, environmental policy, housing and construction, and business investment (Calthorpe and Fulton, 2001; O'Connor *et al.,* 2001), all of which have an impact on population change.

Reimer and Bollman (2006) emphasised the importance of financial capital, human skills and abilities, social capital, and natural resources in their model of Canada's rural economy.

The Australian economy has been transformed massively over the last forty years. In that time, some public assets have been privatised, and businesses have experienced less (or more) regulation. There has been a policy shift as part of the reform including reduction in tariffs and other forms of industry protection in line with World Trade Organisation (WTO) guidelines, the floating of the Australian dollar, and the introduction of National Competition Policy to encourage the more effective use of government assets (Beer and Clower, 2009). A number of studies (e.g., Australian Local Government Association, 1999; Budge, 2006; McKenzie and Frieden, 2010) have focussed on sustaining and developing regional economies, and bodies such as Regional Development of Australia and Regional Development Victoria have helped to implement government policy on these issues. See, for example, the Regional Development Victoria's recent "Ready for Tomorrow: A Blueprint for Regional and Rural Victoria" (Regional Development Victoria, 2010).

Researchers have examined population growth and decline employing a range of *analytical methodologies*. For example, Beer and Clower (2009) examined the impact of economic reforms on population growth and decline in regional cities, and they used cluster and regression analysis to examine the drivers of growth. Their regression analysis shows that increasing economic specialisation results in predictable development.

Other methodologies include the Delphi process (Brown 1968), which involves reliance on judgement of experts and seeks group solutions to a problem or a group estimate of some unknown numerical quantity. Sorensen and Weinand (1991) used factor analysis and Ward's clustering algorithm to study regional wellbeing in Australia and demonstrated complex spatial variations in socio-economic well-being in regional areas. Lee and Carter (1992) used a demographic model and time series to forecast mortality. The method combines demographic model with statistical time series methods and is based on long-term historical patterns and trends. Sanderson (1998) used the Wheeler forecasting model, which uses international data in forecasting fertility. He suggests that Wheeler's model has the simplest demographic structure. These studies concentrate on demographic local dynamics like births, deaths, and migration to track population changes.

Other analysts have used clustering analysis for pattern recognition in larger data sets (see, for example, Freestone *et al.*, 2003; Beer and Clower, 2009; Beer and Maude, 1995). Clustering algorithms can be used to analyse large data sets comprising a myriad of economic, social and demographic variables for numerous samples (SLAs in this study). They seek to group samples with similar characteristics and ensure maximum statistical separation from other contrasting clusters. In this process of pattern recognition, they simplify understanding of those large data sets. Clustering technique using the *k*-means algorithm has been adopted in this study. The *k*-means algorithm provides an alternative clustering technique to clarify population change.

3 OBJECTIVES, DATA AND ANALYTICAL METHODOLOGY

The objective of this study is to address the impact of selected socio-economic factors on population change in regional Australia, using the *k*-means clustering technique because of its recognised ability to analyse large data sets. Building upon previous studies (see, for example, Beer and Clower, 2009; Beer, 1999; Beer and Maude, 1995) this study initially uses *k*-means clustering algorithm to cluster data of variables (Industry of employment, Occupation type, etc) once for each variable type separately. The data set is based on SLA geographical classification. The output of clustering is clusters relevant to each of the above mentioned variables. These generated clusters are then used in a regression model to examine their impact on population change. This stage identifies clusters which are impacting on population growth and decline. In a further analysis each cluster is separately examined to explore how the population change trend inside those clusters relate to the regression analysis results.

The *k*-means algorithm considers each sample (SLAs in this study) in a data set as a point in *n*-dimensional space (R^n) and chooses *k* centres (also called centroids) and assigns each point to the cluster nearest the centre. The centre is the average of all the points in the cluster, that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. The limitation with the *k*-means algorithm is that although it is known to be an efficient clustering algorithm, it is sensitive to the choice of starting points. It can converge on local minima and these local minima may be significantly different from global solutions as the number of clusters increases (Bagirov and Mardaneh, 2006; Bagirov, 2008).

The data used for this paper includes *Census* data for 2001-2006. These were sourced from the Australian Bureau of Statistics (ABS, 2010a), as follows:

- Industry of employment (18 categories)
- Occupation type (8 categories)
- Employment status (6 categories)
- Individual weekly income (12 categories)
- Education level (5 categories), and
- Age group (21 categories)

These are specified in full in Appendix 1.

The study used all industry and occupation categories. Since other variables were too detailed (e.g. individual weekly income with 12 categories; employment status with 6 categories, etc) and that level of detail was not necessary for the analysis, for simplicity those were merged into fewer categories, as specified below:

- the six employment status categories were narrowed down to employed, unemployed, and not in the labour force;
- the 12 income groups were collapsed into five categories – negative income, nil income, \$1- \$999, \$1,000 - \$1,999 and \$2,000 or higher;
- age groups were merged to form three new categories - aged 25-39, 40-64, and older than 64; and
- education levels were merged to form two new categories described here as high level of tertiary and postgraduate, and low level of tertiary and postgraduate.

A database was generated using these ABS data, divided into SLAs (ABS, 2010a).

So far as SLAs were concerned, it was necessary to devise a means to examine only regional (as opposed to metropolitan) regions. This was done by eliminating 'major urban' areas, that is, SLAs with populations exceeding 100,000, and then by eliminating SLAs with fewer than 100,000 people, but on the fringes of major urban areas using the Australian Standard Geographical Classification (ASGC)(See ABS, 2006b and ABS, 2009). This procedure reduced the number of SLAs from 1431 to 726. Then outliers with extreme population growth or decline values were detected. There were 37 of these outliers as specified in Appendix 2. Outliers skew the data and distort the analysis results, so they were eliminated, reducing the number of SLAs from 726 to 689. This group of 689 SLAs were used in the analysis. Table 1 provides a breakdown of these 689 SLAs in terms of categories of population change. Allowance was made for the fact that there had been changes in some SLAs between 2001 and 2006 *Census*.

4 REGIONAL SLA GROWTH AND DECLINE

Population growth and decline rates are presented in Table 1 in six categories encompassing all growth and decline rates. In total, 38% of regional SLAs had a declining population and 62% increasing. Within the *declining* group, 16% of SLAs experienced a decline of more than 5%, while within *increasing* group, 26% of SLAs experienced an increase of up to 5%. Some SLAs have gained population and / or lost population at different rates, and this analysis shows an overall increase of 6% in the population of regional Australia between 2001 and 2006. Beer and Clower (2009, p378) suggest that "overall, the number of regional cities in Australia has grown, partly as a consequence of national population and economic growth".

Table 1. Categories of Population Change (%) for Regional SLAs. Source: the Author.

Categories of change	No of SLAs within each Category	%SLAs within each Category	Population for SLAs within each category		Population change % within each category
			2001	2006	
Decline more than	111	16	289916	262009	-10
Decline 0 to 5%	153	22	1015310	995915	-2
Increase 0 to 5%	178	26	1627238	1667230	2
Increase 5-10%	118	17	1431068	1530253	7
Increase 10-15%	50	8	504809	568220	13
Increase more than	79	11	779915	968299	24
Total	**689**	**100**	**5648256**	**5991925**	**6**

5 ANALYSIS

The analysis to produce the results outlined in this paper followed a two-stage process. In the first stage of analysis, the *k*-means clustering algorithm was used to cluster all SLAs six times, once for each variable type. The clusters generated were then used in the second stage of the analysis, which involved the application of multiple regression analysis.

Depending on the number of categories for each variable, the cluster analysis generated different numbers of clusters ranging from 2 to 7, as follows: industry (seven clusters IndusC1 – IndusC7); occupation type (seven clusters OccupC1 – OccupC7); employment (three clusters EmployC1 – EmployC3); income (seven clusters IncomeC1 – IncomeC7); education level (two clusters EduC1 – EduC2); and age (two clusters AgeC1 – AgeC2).

Any cluster (e.g. IndusC1) relating to each variable (e.g. Industry of employment) includes a combination of categories (e.g. Retail Trade, Public Administration and Safety, Mining, and etc.) These categories each have different mean scores. Under this analytical technique, the category with the highest mean score within a cluster determines the type of that cluster. For example, in cluster IndusC1, Retail Trade has the highest mean score (12.20), and the lowest mean score (3.22) is for the Wholesale Trade category. Therefore the entire cluster is considered to be represented by Retail Trade. Amongst the selected categories the highest mean score (53.20) belongs to Public Administration and Safety. This indicates that, on average, the proportion of employed people in Public Administration and Safety industry is higher than all the other industry types combined. The category of Agriculture, Forestry and Fishing appears to have the highest mean score within two clusters (IndusC6 and IndusC7). Cluster descriptions and the highest mean scores for each category are presented in Table 2.

Analysis shows that for individual weekly income when the number of clusters is seven, data is represented better. For this reason the number of clusters (7) exceeds the total number of categories (5). In this case, clusters are only reported if they include at least one category with the highest mean score. For Employment status, since the Employed category is so dominant (appears with the highest mean value) within all three relevant clusters (meaning that majority of people belong to this category), the values of the other categories are repressed within all clusters. To overcome this problem, two categories ('Unemployed' and 'Not in labour force') with higher mean scores, from the same cluster have been selected. For this reason the cluster (EmployC3) has been reported twice in Table 2.

As with industry of employment, cluster analysis revealed seven clusters for the occupation type. Each occupation type cluster includes a combination of different occupation categories with different mean scores. In this case, the highest mean score belongs to Managers within cluster C6 (OccupC6). This indicates that the proportion of employed people in the Managers category is higher than all the other occupation types. The Managers category and the Labourers category both happen to have the highest mean score in two clusters (OccupC4 and OccupC5 for Labourers, OccupC3 and OccupC6 for Managers).

With the other categories: for employment status, the highest mean score (47.11) belongs to the Employed category; for individual weekly income, seven clusters emerged, with the highest mean score belonging to the \$1-\$999 category within cluster 7 (IncomeC7); for education, the highest mean score relates to the High level of tertiary and postgraduate category (cluster EduC2, where the proportion of people having high rather than low level of tertiary and postgraduate education); for age group, two clusters emerged, with the highest mean score belonging to the 40-64 years old category. Cluster C1 (AgeC1) appears to include the highest mean score for both the 40-64 years old and 65 and older categories.

Variables such as gender, state of residence, etc. (as non-continuous variables) are encountered frequently in the social sciences and are considered as “nominal-level” variables. At this stage and as a result of the above process, categories with the highest mean score are selected (now each representing a cluster e.g. IndusC1, Age2) and are summarised in Table 2. These clusters (as nominal-level dummy variables), not the master variables (e.g. Industry of employment, Age group), were then subjected to multiple regression and were included in the regression model. In other words these clusters (as independent variables) were used to examine their impact on population change (as dependent variable).

Table 2. Variables and Categories, Clusters and Mean Values for Each Cluster. Source: (ABS, 2006a)

Clusters	Categories (Name)	Highest Mean Score (%)/ Cluster
Industry of employment (7)		
IndusC1	Retail Trade	12.20
IndusC2	Public Administration and Safety	53.20
IndusC3	Accommodation and Food Services	23.75
IndusC4	Professional, Scientific and Technical Services	13.12
IndusC5	Mining	30.98
IndusC6	Agriculture, Forestry and Fishing	23.01
IndusC7	Agriculture, Forestry and Fishing	46.41
Occupation type (7)		
OccupC1	Professionals	28.75
OccupC2	Technicians and Trades Workers	16.75
OccupC3	Managers	24.20
OccupC4	Labourers	37.98
OccupC5	Labourers	37.17
OccupC6	Managers	41.13
OccupC7	Machinery Operators And Drivers	18.57
Employment status (3)		
EmployC1	Employed	47.11
EmployC3	Unemployed	2.79
EmployC3	Not in labour force	29.32
Individual weekly income (7)		
IncomeC2	Nil income	6.39
IncomeC5	Negative income	1.91
IncomeC6	$1000- $1999	18.48
IncomeC6	$2000 and more	8.83
IncomeC7	$1- $ 999	56.56
Education level (2)		
EduC2	High level of tertiary and postgraduate	29.72
Age group (2)		
AgeC1	40-64 years old	34.25
AgeC1	65 and older	14.47
AgeC2	25-39 years old	22.29

Initial regression analysis was conducted by including each set of clusters (industry of employment, etc) in a model. This analysis revealed that neither occupation type nor employment status had a statistically significant impact on population change and were not therefore included in the analysis. Their importance may have been subsumed under industry of employment categories. The regression model therefore includes dummy variables associated with the industry of employment, individual weekly income, education level and age group:

Population change = $a + b_1$ IndusC1 + b_2 IndusC2 + b_3 IndusC3 + b_4 IndusC4 + b_5 IndusC5 + b_6 IndusC7 + b_7 IncomeC1 + b_8 IncomeC2 + b_9 IncomeC3 + b_{10} IncomeC4 + b_{11} IncomeC6 + b_{12} IncomeC7 + b_{13} EduC2 + b_{14} AgeC2

The standard ordinary least-square (OLS) regression method was used for this analysis.

Since the population change data was not normally distributed, it was therefore transformed using a log (10) of the percentage change between 2001 and 2006 population data. After transformation some cases appeared to have negative values, and therefore a constant was added and applied across whole cases of the population change variable to maintain positive scaling. Results for the multiple regression analysis are shown in Table 3.

6. RESULTS AND DISCUSSION

"Goodness of fit" measure (R^2) indicates the overall fit between the model and the data. (For details see Gow, 2007). For this study, regression analysis yielded a relatively low R^2 (R^2 =0.35). The R^2 value indicates that the independent variables used in this analysis were able to explain up to about 35% of the variation in population change data. This low R^2 reflects the nature of the analysis in which there are many "other" variables - not included in the present analysis (such as communication, transportation, health provision, proximity to the metropolitan centre, size of the regional centre, etc.) - that also explain population change.

It was not intended that this study would provide a comprehensive account of "what" influences population growth and decline, but instead it aimed to identify some of the variables that "significantly" account for population growth and decline and could help to explain patterns of population change. The variables used for this analysis are not perfect indicators of the constructs the study intended to measure. All measures are subject to some error and mean-based variables such as income do not necessarily reflect the split between high and low income within a region. Any measurement error would tend to alter true relationships, and thus the results could alter the true size of the effects being measured (leading to a low R^2).

Under each variable (industry of employment, individual weekly income, education level, age group) one of the clusters, with the lowest coefficient value (associated with the lowest population growth) constitutes the 'base' (reference) cluster and all the other clusters (of a particular variable) are compared against this base cluster. As a rule when incorporating these clusters (as dummy variables) into a regression model, only *n*-1(where *n* signifies the number of clusters) clusters (variables) are entered to represent the required information. As a result, these base clusters (IndusC6; IncomeC5; EduC1; AgeC1) do not appear in Table 3. Comparison of other clusters to the base cluster helps to explain whether other clusters contribute to population growth or not, and if they do contribute what is the significance of their contribution compared to the base (lowest) cluster.

Nonetheless, the analysis has produced results that provide a strong indication of some of the variables that have an impact on population growth and decline in regional Australia. Overall the analysis indicates that generally clusters with a positive regression coefficient show a higher rate of population gain. Table 3 provides appropriate summary information of statistical results, and the proportion of SLAs inside each cluster that are facing population growth or decline.

Table 3. Model Regression Results. Source: the Author.

Independent Variables	Categories (Name)	Coefficient	SE coefficient	Standardised coefficient	T-Stat.	P-value	% of members (SLAs) inside each cluster facing negative or positive population growth		Net % Change inside each cluster
							Negative	Positive	
Intercept		0.660	0.009		72.161	0.001			
Industry of employment									
IndusC1	Retail Trade	**0.026**	0.006	0.178	4.372	0.001	28	72	6
IndusC2	Public Administration and Safety	-0.014	0.012	-0.054	-1.154	0.249	48	52	-1
IndusC3	Accommodation and Food Services	**0.052**	0.019	0.09	2.781	0.006	29	71	5
IndusC4	Professional, Scientific and Technical Services	**0.049**	0.015	0.125	3.176	0.002	10	90	19
IndusC5	Mining	**-0.100**	0.014	-0.296	-7.358	0.001	53	47	-2
IndusC7	Agriculture, Forestry and Fishing	**-0.034**	0.008	-0.181	-4.332	0.001	62	38	-1
Individual weekly income									
IncomeC1	$1- $ 999	0.007	0.01	0.045	0.720	0.472	28	72	7
IncomeC2	Nil income	0.006	0.015	0.017	0.378	0.706	43	57	3
IncomeC3	$1- $ 999	**0.026**	0.014	0.094	1.909	0.057	27	73	11
IncomeC4	$1- $ 999	0.021	0.014	0.059	1.468	0.142	19	81	6
IncomeC6	$1000-$1999 & $2000 and more	**0.136**	0.019	0.323	7.227	0.001	3	97	36
IncomeC7	$1- $ 999	-0.002	0.009	-0.014	-0.210	0.834	40	60	3
Education level									
EduC2	High level of tertiary and postgraduate	**0.014**	0.006	0.097	2.429	0.015	15	85	24
Age group									
AgeC2	25-39 years old	**0.022**	0.007	0.14	3.040	0.002	35	65	12

Notes: R^2=0.353; adjusted R^2=0.340; SEE=0.056; Coefficients in bold: Significant at the 95% level.

Industry of Employment

Three industry clusters were positively associated with population growth and therefore linked with attracting population to regional areas: Accommodation and Food Services (IndusC3), Professional, Scientific and Technical Services (IndusC4), and Retail Trade (IndusC1), and two produced positive associations with population decline: Agriculture, Forestry and Fishing (IndusC7), and Mining (IndusC5).

Accommodation and Food Services (IndusC3) has the highest positive coefficient of 0.052 and a t-value of 2.78. There is a positive correlation with the 25-39 years old age group ($r= 0.15$, $p<0.001$) and the High level of tertiary and postgraduate education cluster ($r= 0.39$, $p<0.001$). This indicates that Accommodation and Food Services industry attracts younger people with higher education, which could be a temporary attraction. There is also a positive correlation between Accommodation and Food Services with the $1000-$1999 income level ($r= 0.18$, $p<0.001$). Within cluster IndusC3, 71% of SLAs are facing population increase as opposed to 29% of which are losing population.

Professional, Scientific and Technical Services (IndusC4) produced the second-highest positive coefficient. The coefficient for this cluster is 0.049 with a t-value of 3.17. The findings indicate a positive correlation with the 25-39 years old age group ($r= 0.28$, $p<0.001$), the High level of tertiary and postgraduate education cluster ($r= 0.71$, $p<0.001$), and the $2000 and more income level group ($r= 0.44$, $p<0.001$). As was the case with Accommodation and Food Services, this industry is labour intensive and it absorbs highly educated people. The IndusC4 cluster has 90% of its SLAs showing an increase and 10% losing population.

Retail Trade (IndusC1) follows the pattern of the two previously-mentioned industries. Although this cluster does not have as large a positive coefficient (0.026) as the other two clusters (IndusC3, IndusC4), the t-value of 4.37 does indicate the relative importance of this cluster. Retail Trade also shows a positive correlation with the High level of tertiary and postgraduate ($r= 0.46$, $p<0.001$) however, it correlates with the 40-64 years old ($r= 0.23$, $p<0.001$) and the 65 and older age groups ($r= 0.49$, $p<0.001$). Similarly there is a positive correlation between Retail Trade and the $1000-$1999 income level ($r= 0.12$, $p<0.001$). Within Retail Trade (IndusC1), 72% of SLAs are increasing in population and 28% are losing.

Agriculture, Forestry and Fishing (IndusC6) constitutes the base cluster, and all other industry clusters have been compared with this cluster. Another Agriculture, Forestry and Fishing cluster (IndusC7) shows a negative coefficient of -0.034 and a negative t-value of -4.33. There is a negative correlation between *Agriculture, Forestry and Fishing* with both the 25-39 years old age group ($r= -0.03$, $p<0.352$) and the High level of tertiary and postgraduate education cluster ($r=-0.05$, $p<0.174$), but neither are statistically significant. This indicates that on the one hand the industry is not labour intensive and on the other hand although highly educated people are not absorbed in this sector directly, which could mean they were absorbed in Professional, Scientific and Technical Services industry which, indirectly, provides services to Agriculture, Forestry and Fishing. This industry is a commodity export industry and has a negative impact on population growth. Industrial restructuring and the need for improvement in efficiency was identified by the Commonwealth Department of Transport and Regional Services (2001). Within the Agriculture, Forestry and Fishing cluster (IndusC7) which shows a negative coefficient in Table 3, up to 62% of SLAs are facing population loss and only 38% are gaining population. This indicates that clusters with a negative coefficient value have a high rate of population loss.

Similarly the *Mining* industry (IndusC5) has a negative impact on population growth with a negative coefficient of -0.100 and a t-value of -7.36. The analysis also suggests that the correlation between Mining and the 25-39 years age group is positive but this correlation is small ($r= 0.28$, $p<0.001$). Similarly the correlation between Mining and the High level of tertiary and postgraduate ($r= 0.01$, $p<0.669$) is positive but not statistically significant. A positive correlation also exists between Mining and the $2000 and more income level ($r= 0.57$, $p<0.001$). This indicates that on the one hand the industry is relatively labour intensive however it does not attract numerous highly educated people. On the other hand the income level attracts people to the industry. For Mining (IndusC5), 53% of SLAs are losing population and 47% are gaining population. The Mining cluster shows a negative coefficient in Table 3.

As is evident, clusters with a higher positive coefficient appear to have a higher rate of positive population change. This supports the work of Beer and Clower (2009) based on which regional cities with an employment structure focused on entertainment, tourism and leisure industries were identified for the first time in 1991. Similarly an agribusiness cluster and remote tourism grouping emerged as growth areas for the first time in 2001. Beer (1999) suggests that most of the population growth in regional areas is related to tourism, recreation and leisure industries. The findings of this current study add to the importance of the role of ‘industry of employment’ in development of the regions.

In emphasising the role of industry, Beer and Clower (2009) suggest that if a region or city specialises in industries with limited prospects they may stagnate, whilst areas that link their economy to a growing industry will benefit into the future. Beer *et al*., (2003) suggest that the mix within industries is important as well. This was identified by O’Connor *et al*., (2001) which suggest that attraction of New South Wales and Victoria to businesses could be seen as a result of concentration of particular types of industries in those areas. On the other hand similar to the Agriculture, Forestry and Fishing the Mining industry is a commodity export industry and had a negative impact on population growth between 2001 and 2006. Clusters with mining concentration experience population loss despite the success in mining exports as identified by Beer and Clower (2009).

Income

The highest coefficient (0.136) belongs to “$1000-$1999 and $2000 and more” income level (IncomeC6). Incomes of $1000-$1999 are strongly associated with Accommodation and Food Services and the Retail Trade sectors. Similarly incomes of $2000 and more are strongly associated with Professional, Scientific and Technical Services, and the Mining industries. Analysis of individual weekly income reveals that the negative income (IncomeC5) is directly associated with population decline. On the other hand, weekly incomes over $1000 have a positive impact on population change.

Within the “$1000- $1999 and $2000 and more” income levels (IncomeC6) 97% of SLAs are gaining population and only 3% are losing population. For the $1- $999 income level (Income C1, C3, C4, C7) the average rate is 71.5% gain and 28.5% loss.

Education

The High level of tertiary and postgraduate (EduC2) has a positive coefficient of 0.014 and a t-value of 2.42. High level of tertiary and postgraduate also shows a strong correlation ($r= 0.63$, $p<0.001$) with the $1000-$1999 as well as the $2000 and more income levels ($r=0.47$, $p<0.001$). High level of tertiary and postgraduate has the lowest correlation with the $1-$999 income level ($r=0.30$, $p<0.001$), but the correlation is still significant. This shows that a higher level of education contributes positively to higher population growth. In contrast, a lower level of education does not have as much impact on population growth. Within the High level of tertiary and postgraduate cluster (EduC2), 85% of SLAs are gaining population and 15% are losing population.

Age

Analysis indicates that the 25-39 years old (AgeC2) age group has a positive coefficient of 0.022 and a t-value of 3.04. Not surprisingly the 25-39 years old age group has a positive correlation ($r= 0.20$, $p<0.001$) with the High level of tertiary and postgraduate education. Analysis of age groups also revealed that the age groups of 40-64 and 65 and older do not have as much impact on population growth. This is relevant as different age groups have different impact on population change. Ageing patterns are crucial in understanding how to sustain population in regional Australia. Based on this analysis, for the 25-39 years old age group (AgeC2) 65% of SLAs are gaining population and 35% are losing. Age categories in regional Australia appear to be skewed compared to the urban capital centres.

Regional areas with the capacity to offer some services such as education have population growth potential as identified by Beer (1999) and this is clearly supported by the findings of this current study. Attracting skilled population to regional Australia is a very challenging task and every regional area needs to be considered separately based on their different needs and resources. Policy initiatives such as investing in education in regional areas are most likely to help with the development of such areas, as identified by Regional Development Australia (2010) and supported by the findings of this study. Smith *et al.*, (2011) reinforce this in their investigation of innovative capacity building by investing in tertiary education, and learning and development systems, in order to support economic development. Such investments in education are particularly pertinent to regional areas which suffer from lack of services and critical mass, as noted by Regional Development Australia (2010).

7. CONCLUSION

This paper builds upon studies by Smith (1965), Beer (1999), Beer and Maude (1995), Beer and Clower (2009), and Freestone *et al.*, (2003), but takes a different approach in order to investigate the impact of socio-economic factors on population change in regional Australia. This study used a different clustering algorithm (*k*-means) to cluster SLAs in regional Australia to investigate the impact of emerging clusters on population growth and decline. The study further examined each cluster separately to investigate whether the role of emerging clusters on population growth and decline are linked to the population change trend inside each cluster. In summary, the clusters that contribute more strongly to population growth are industry clusters Accommodation and Food Services, Professional, Scientific and Technical Services, and the Retail Trade industries, in association with the 25-39 years old age group, and individual weekly income of more than $1000, and a High level of tertiary and postgraduate education. Agriculture, Forestry and Fishing as well as the Mining industries show a positive association with population decline. This could be due to the fact that recent changes, particularly the mining boom is not captured in 2001-2006 data and could be reflected in 2011 *Census* data. These findings suggest that investment policies and the investment in particular industry types in regional areas should be reviewed.

This paper used two distinct methodologies to investigate the link between the impact of clusters on population change and the population change trend inside each cluster. The results show that clusters with a positive impact are clusters with a higher rate of population gain, while clusters with a negative impact are clusters that show a higher rate of population loss.

Based on the results outlined in this paper, further research adopting an optimisation approach and using the *k*-means clustering algorithm would create opportunities for further comparative analysis of the population change in regional and metropolitan Australia. Such clustering research could identify further trends that have policy implications for addressing regional needs on a more specialised and specific basis for both regional and metropolitan areas separately.

REFERENCES

Adelman, I. (1963). An Econometric Analysis of Population growth. *The American Economic Review*, 53(3), pp. 314-339.

Australian Bureau of Statistics. (2006a). *Census data.*

Australian Bureau of Statistics. (2006b). Statistical Geography Volume1- Australian Standard Geographical Classification (ASGC), *ABS catalogue No. 1216.0*.

Australian Bureau of Statistics. (2009). Australian Standard Geographical Classification (ASGC), *ABS catalogue No. 1216.0*.

Australian Bureau of Statistics. (2010a). *Census CDATA online*, 2006.

Australian Bureau of Statistics. (2010b). *Regional Population Growth, Australia 2008-09. ABS catalogue No. 3218.0.*

Australian Government Department of Regional Australia, Regional Development and Local Government. (2010). *Homepage*, Retrieved September 25, 2010, from: http://www.regional.gov.au/regional/

Australian Local Government Association. (1999). *Opportunities for local government, Competitive Regions*, Retrieved September 25, 2010, from: www.alga.asn.au/publications/Developing_competitive_regions.pdf

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

Bagirov, A. M. and Mardaneh, K. (2006). Modified global *k*-means Algorithm for Clustering in Gene Expression Data Sets. *Intelligent systems for Bioinformatics2006*, Hobart, Australia, Australian Computer Society (ACS).

Beer, A. (1999). Regional Cities Within Australia's Evolving Urban System, 1991-96. *Australasian Journal of Regional Studies*, 5(3), pp. 329-348.

Beer, A. and Clower, T. (2009). Specialisation and Growth: Evidence from Australia's Regional Cities. *Urban Studies*, 46, pp. 369-388.

Beer, A. and Maude, A. (1995), Regional Cities in The Australian Urban system, 1961-1991. *Urban Policy and Research*, 13(3), pp. 135-148.

Beer, A. and Keane, R. (2000). Population decline and service provision in regional Australia: A South Australian Case Study, *People and Place*, 8(2), pp. 69-75.

Beer, A., Maude, A. and Pritchard, B. (2003). *Developing Australia's Regions: theory and practice*, University of New South Wales Press, Sydney.

Brown, B. B. (1968). *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*, The Rand Corporation, California.

Budge, T. (2006). Sponge Cities and Small Towns: a New Economic Partnership. In M. Rogers and D. R. Jones (Eds), *The Changing Nature of Australia's Country Towns*, VURRN press, Ballarat.

Calthorpe, P. and Fulton, W. (2001). *The regional City: planning for the end of sprawl*, Washington DC, Island press.

Courvisanos, J. (2009). Innovation policy and social learning: an economic framework for sustainable development in regional Australia. In Martin, J., Rogers, M. and Winter, C (Eds), *Climate Change in Regional Australia: Social Learning and Adaptation*, VURRN Press, Ballarat.

Commonwealth Department of Transport and Regional Services. (2001). *The Success Factors-managing change in regional and rural Australia*, Report commissioned by Regional Women's Advisory Council, pp. 5-10.

Freestone, R., Murphy, P., and Jenner, A. (2003). The functions of Australian towns, revisited. *Tijdschrift voor Economische en Sociale Geografie,* 94 (2), pp. 188-204.

Frisbie,W. P. and Poston. Jr., D.L. (1975). Components of Sustenance Organization and Nonmetropolitan Population Change: a Human Ecological Investigation. *American Sociological review*, 40(6), pp. 773-784.

Frisbie,W. P. and Poston. Jr., D.L. (1978). Sustenance Differentiation and Population Redistribution. *Social Forces*, 57(1), pp. 42-56.

Goetz, S. J. and Debertin, D. L. (1996). Rural Population Decline in the 1980s: Impacts of Farm Structure and Federal Farm Programs. *Amer J Agr Econ*, 78, pp. 517-529.

Gow, D. J. (2007). *Fundamentals of multiple regression analysis*, 2010 ACSPRI summer course lecture notes, The Australian National University, Canberra.

Infrastructure Australia. (2010). *Major Cities Unit: Population and Settlement*, Chapter 3. pp. 27-48.

Lee, R. D. and Carter, L. R. (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), pp. 659-671.

McKenzie, F. and Frieden, J. (2010). *Regional Victoria: Trends and Prospects*, Melbourne, Victoria: Victorian Government, Department of Planning and Community Development. Retrieved September 25, 2010, from: www.dpcd.vic.gov.au

Millward, H. (2005). Rural Population Change in Nova Scotia, 1991-2001: bivariate and multivariate analysis of key drivers. *The Canadian Geographer*, 49 (2), pp. 180-197.

O'Connor, K., Stimson, R. and Daly, M. (2001). *Australia's Changing Economic Geography: A society Dividing*, Oxford University Press, Melbourne.

Polese, M. and Shearmur, R. (2006). Why some regions will decline: A Canadian case study with thoughts on local development strategies. *Papers in Regional Science*, 85 (1), 23-46.

Regional Development Australia. (2010). *Strategies for Regional Growth*, Regional Development Australia Factsheet No.1.

Regional Development Victoria. (2010). *Ready for Tomorrow: A Blueprint for Regional and Rural Victoria*, Retrieved September 25, 2010, from: www.rdv.vic.gov.au

Reimer, B. and Bollman, R. D. (2006). The New Rural Economy: Key Observations for Research and Policy in the Canadian Context. In M. Rogers and D. R. Jones (Eds), *The Changing Nature of Australia's Country Towns*, VURRN press, Ballarat.

Sanderson, W. C. (1998). Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models. *Population and Development Review*, 24, pp. 88-117.

Shumway, J. M. and Davis, J. A. (1996). Nonmetropolitan Population Change in the Mountain West: 1970-1995. *Rural Sociology*, 61(3), pp. 513-529.

Smith, A., Courvisanos, J., Tuck, J., and McEachern, S. (2011). Building innovation capacity: the role of human capital formation in enterprises. In Curtin, P., Stanwick, J. and Beddie, F (Eds), *Fostering Enterprise: The Innovation and Skills Nexus – Research Readings*. NCVER, Adelaide.

Smith, R. H.T. (1965). Method and Purpose in Functional Town Classification. *Annals of the Association of American Geographers, 55 (3)*, pp. 539-548.

Sorensen, T. and Weinand, H. (1991). Regional Well-Being in Australia Revisited. *Australian Geographical Studies*, 29 (1), pp. 42-70.

Standing Committee on Regional Development for the Regional Development Council. (2010). *Attracting and Retaining Skilled People in Regional Australia*, 1-55.

Appendix 1. Socio-economic Factors for Analysis. Source: the Author.

Industry of employment (18 categories)	**Occupation type (8 categories)**	**Employment status (6 categories)**
Agriculture, Forestry and Fishing; Mining; Manufacturing; Electricity, Gas, Water and Waster Services; Construction; Wholesale Trade; Retail Trade; Accommodation and Food Services; Transport, Postal and Warehousing; Information, Media and Telecommunications; Financial and Insurance Services; Rental, Hiring and Real Estate Services; Professional, Scientific and Technical Services; Administrative and Support Services; Public Administration and Safety; Education and Training; Health Care and Social Assistance; Arts and Recreation Services; Other Services. All industry categories were used	Managers; Professionals; Technicians and Trades Workers; Community and Personal Service Workers; Clerical and Administrative Workers; Sales Workers; Machinery Operators and Drivers; Labourers. All occupation types were used	Employed, worked full-time; Employed, worked part-time; Employed, away from work; Unemployed looking for full-time work; Unemployed looking for part-time work; Not in the labour force. Collapsed to: • Employed • Unemployed • Not in the labour force
Individual weekly income (12 categories)	**Education level (5 categories)**	**Age group (21 categories)**
Negative income Nil income Weekly income in brackets of: $1-$149; $150-$349 (sets of $100); $350-$799 (sets of $150); $800-$1399 (sets of $200); $1400-$1999 (sets of $300); $2000-$3999 (sets of $500); $4000 or more. Collapsed to: • Negative income • Nil income • $1 - $999 • $1000 - $1999 • > $1999	Postgraduate Degree Level; Graduate Diploma and Graduate Certificate Level; Bachelor Degree Level; Advanced Diploma and Diploma Level; Certificate Level. Collapsed to: • High level of tertiary and postgraduate • Low level of tertiary and postgraduate	In sets of 5 years as: 0-4 years; 5-9 years; …; 95-99 years; 100 and over. Collapsed to: • 25-39 • 40-64 • 65 and older

Appendix 2. SLAs with High Rate of Population Growth or Decline. Source: the Author

SLA	Population 2001	Population 2006	%Change
VIC			
Melton (S) - East	16091	40776	153.4
Cardinia (S) - Pakenham	17990	28457	58.2
Gr. Bendigo (C) - S'saye	5158	6730	30.5
Mitchell (S) - South	16954	21014	23.9
QLD			
Bowen Hills	900	1652	83.6
City - Inner	1021	2823	176.5
City - Remainder	1827	4658	155.0
Fortitude Valley	3106	5673	82.6
Kangaroo Point	5667	7235	27.7
Newstead	2878	5113	77.7
Spring Hill	3485	5483	57.3
Caloundra (C) - Caloundra N.	18398	22493	22.3
Caloundra (C) - Caloundra S.	15778	21342	35.3
Buderim	33178	42734	28.8
Hervey Bay (C) - Pt A	39599	50864	28.4
Cambooya (S) - Pt A	3241	4143	27.8
Cambooya (S) - Pt B	3241	4143	27.8
Crow's Nest (S) - Pt A	6450	9205	42.7
Jondaryan (S) - Pt A	5646	7560	33.9
Kirwan	20096	24683	22.8
Fitzroy (S) - Pt A	5003	6542	30.8
Nebo (S)	2094	2673	27.7
City	1627	2167	33.2
Douglas	3346	6436	92.3
Mt Louisa-Mt St John-Bohle	4125	5881	42.6
Oonoonba-Idalia-Cluden	1975	3749	89.8
Weipa (T)	2173	3007	38.4
NSW			
Palerang (A) - Pt A	7833	9567	22.1
SA			
Unincorp. Far North	3335	1644	-50.7
WA			
Capel (S) - Pt A	2859	6319	121.0
Laverton (S)	1202	786	-34.6
Menzies (S)	349	238	-31.8
Ravensthorpe (S)	1504	2062	37.1
Dardanup (S) - Pt A	6418	8158	27.1
East Pilbara (S)	5628	7007	24.5
Chittering (S)	2936	3700	26.0
NT			
Thamarrurr (CGC)	1665	2222	33.5