THE ROLE OF EDUCATION IN REGIONAL INCOME DETERMINATION- A CROSS SECTIONAL STUDY OF SMALL AREAS IN QUEENSLAND[1][2]

Bernard Trendle
Labour Market Research Unit, Department of Employment and Training, Locked Mail Bag 527, GPO Brisbane, QLD 4001.

Warren Pears
Labour Market Research Unit, Department of Employment and Training, Locked Mail Bag 527, GPO Brisbane, QLD 4001.

ABSTRACT: This study explores cross regional variation of income within Queensland. This is done using data from the 2001 Census of Population and Housing conducted by the Australian Bureau of Statistics. Of particular interest is the role that education has played in regional income determination. The study commences with an analysis of the spatial pattern of regional income levels in Queensland. A formal model that includes a role for education in determining differences in income levels is developed and the census data is used to test the validity of the model. It is found that education is a significant determinant of the cross regional variation in income levels, with both higher education and vocational training playing a role in this difference.

1. INTRODUCTION

Within the field of regional science a large amount of work has been undertaken on regional income. Numerous studies have explored the cross regional variation in income inequality with several aspects of this being explored. For example, Aigner and Heins (1967) explore the role the level of development has had on regional income inequality, while Levernier, Rickman and Partridge (1995) have explored the role of industrial structure. Other studies have explored the role of city size (see, for example, Long, *et al.*, 1977; Garofalo and Fogarty, 1979 and Nord, 1980) and, in the case of Australia, the role of isolation (Maxwell and Peter, 1988).

Education has been found to improve the productivity of labour, for example, Black and Lynch (1996) using survey data from 3,358 U.S. firms, find that human capital is an important determinant of establishment productivity, with the average education level of establishments having a positive and significant effect on firm performance. Similarly, Mason and van Ark (1994), using a matched sample of firms from the Netherlands and UK, find that engineering

---

[1] This paper was presented to the Australian and New Zealand Regional Science Association International (ANZRSAI) Conference, Wollongong NSW, Sept-Oct 2004.

[2] The views expressed in this paper are those of the author and should not be regarded as representing the views of the Queensland Government nor the Department of Employment and Training.

firms in the Netherlands enjoyed higher levels of productivity than their UK counterparts, in part due to their higher levels of workforce skills attributed to the widespread provision of vocational education and training within the Netherlands. From this it can be seen that education is linked to the variation in productivity, a significant determinant of income. For this reason, it would be expected that different levels of education across regions would result in variations in regional income levels.

Within the regional growth literature, education has received considerable attention as a determinant of cross regional variation in economic performance. In this literature Lucas (1998) has demonstrated the importance of human capital in the growth process, while Chatterji (1998) looked at both secondary and higher education and their contribution to cross regional variation in economic growth. Chatterji (1998) found that both tertiary and secondary education were significant determinants of growth in GDP, with tertiary education having a greater influence than secondary education.

While variation of income inequality has been heavily investigated, relatively little work has been done on the determinants of cross regional variation in income. Some examples of this work include Reza (1974), Shah and Walker (1983), Topel (1986) and Topel (1994) for the US, and Molho (1992) for the UK. Reza (1974) examined the relationship between unemployment rates and local wages in U.S. metropolitan areas. In contrast, Shah and Walker (1983) use unit record data to show that money earnings and real incomes vary considerably across regions, even when other factors such as education and industry mix are held constant. Topel (1986) focused on the role of changing local labour market conditions in affecting the geographic distribution of employment and wages, while Topel (1994) used regional differences in the evolution of wage inequality to provide new evidence on the determinants of relative wages. Finally, Molho (1992) provided a review of much of the literature on local pay differentials.

The aim of this study is to determine the variables that have influenced regional income levels within Queensland. This is done using data from the 2001 Census of Population and Housing for the 125 Local Government Areas (LGAs) of Queensland. The following section commences the analysis with an exploration of the spatial variation in regional income across the 125 LGAs of Queensland. This section uses techniques from the rapidly developing field of spatial data analysis. Following this, a formal model explaining the observed cross regional variation in regional income is specified, while section 4 outlines the econometric techniques that will be used in this analysis. Section 5 presents the estimation results and is followed by a brief conclusion.

2. REGIONAL VARIATION IN QUEENSLAND'S INCOME

To examine the determinants average regional income levels in Queensland, we examine cross sectional data collected from the Australian Bureau of Statistics, 2001 Census of Population and Housing at the LGA level. Analysis at this level of spatial geography raises a number of issues, particularly in relation to what is known as the Modifiable Areal Unit Problem. This problem consists of two, interrelated issues, firstly, there are issues regarding the spatial scale of the phenomenon being studied, and secondly, there are issues around the interpretation of results from spatial analysis. The first of these problems may arise if there is a mismatch between the scale of the phenomenon being studied and the spatial units of analysis. This can be addressed by testing for spatial autocorrelation and applying appropriate techniques if it is found to exist. The second problem is frequently called the problem of *ecological fallacy* and follows from the uncertainty of choosing aggregated zonal units. Different units of geography may produce different results, making generalisations difficult. In the current study, we have followed Messner and Anselin (2004) and based our analysis at the spatial scale for which policy may be designed.

The data used in this study are provided in distinct income bands and for this reason a methodology was required to derive average income. In this analysis the proposal of Needleman (1978), which is designed to derive an estimate of the Gini coefficient of regional income inequality was used. This methodology produces an estimate of average regional income as a by-product of the estimation process and so proved useful in this analysis. The average weekly income across the state ranged from $243 in Aurukun Shire to $1,397 in Belyando Shire, a range of $1,154. Aurukun Shire has a high percentage of indigenous persons making up its population, while Belyando Shire is sparsely populated, consisting mainly of mining towns. The average weekly income of all LGAs was $478, with the top quintile having an average of $692, while the average of the bottom quintile was $348.

Table 1 presents selected characteristics of LGAs in the top and bottom quintile. Some noticeable differences can be seen. Average income in the top quintile is twice as high as the bottom quintile, while population is almost four times higher. In the top quintile the proportion of the population with a bachelor or higher degree and the proportion of the population with a certificate level qualification were greater than in the bottom quintile. Similarly, the proportion of the working population in agriculture and in mining was greater in the top quintile as was the proportion of youth (persons aged 15 to 24) in the local population. In contrast, the bottom quintile had a greater proportion of females in the labour force, indigenous persons, unemployed persons and mature-aged persons.

Figure 1 displays a thematic map of Queensland depicting the quintiles of mean income for each of the LGAs. The highest quintile members are predominantly located in mining areas of Central Queensland or in western areas where agriculture is the dominant industry. This group also includes the capital, Brisbane City.

In contrast, the lowest quintile areas occurred in a number of LGAs in the Wide Bay - Burnett region just north of Brisbane. LGAs in the Wide Bay - Burnett region generally have the highest unemployment rate of any region in Queensland and the lowest labour force participation rate, both of which would contribute to low income levels. Other areas of low income include the shires of Mornington, Aurukun and Torres in the Far North of the state, which have significant indigenous populations, Mt Morgan Shire in Central Queensland which has the highest unemployment rate of any LGA in Queensland, Herberton and Hinchinbrook Shires in the Far North and Stanthorpe and Inglewood Shires on the New South Wales border.

Table 1. Characteristics of Regions with Income in the Upper and Lower Quintile, Queensland.

	Top quintile	Bottom quintile	Difference
Income	692.20	348.10	344.10
Population	38,915	10,450	28,465
Proportion of women in the labour force	0.383	0.416	-0.033
Proportion of indigenous population	0.068	0.144	-0.077
Proportion of population with bachelor degree or higher	0.066	0.049	0.017
Proportion of population with a certificate level qualification	0.189	0.165	0.024
Proportion of working population employed in agriculture	0.344	0.204	0.139
Proportion of working population employed in mining	0.123	0.014	0.109
Unemployment rate	0.036	0.107	-0.071
Proportion of population aged between 15 & 24	0.169	0.152	0.016
Proportion of population aged 45 & over	0.391	0.510	-0.119

The data presented in the quintile map suggest that there is some spatial dependence in the distribution of average weekly incomes across the LGAs, i.e. that average weekly income is not randomly distributed across geographic space but some spatial clustering occurs. One way of testing if this clustering is significant is through the Moran scatter plot and Moran I statistic.

Figure 2 provides a Moran scatter plot. Here each LGAs difference from the average value of average income is plotted against the spatial lags or weighted average of the average income of neighbouring local areas. The graph is split into four quadrants. Quadrant 1 represents LGAs with low average income surrounded by high income neighbours; Quadrant 2 has high average income with high average income neighbours; Quadrant 3 has low average income with low average income neighbours and Quadrant 4, high average income with low average income neighbours.

The points in the figure suggest a clustering in quadrant 3 and to a lesser extent quadrant 2, i.e. low average income areas are adjacent to other low average income areas and high average income areas are adjacent to other high average income areas. The trend line through the scatter plot suggests that the spatial clustering is significant and that the regional variation of average income is not randomly distributed throughout the state. The significance of spatial dependence is confirmed using the Moran I test, the results of which are shown at Table 2.

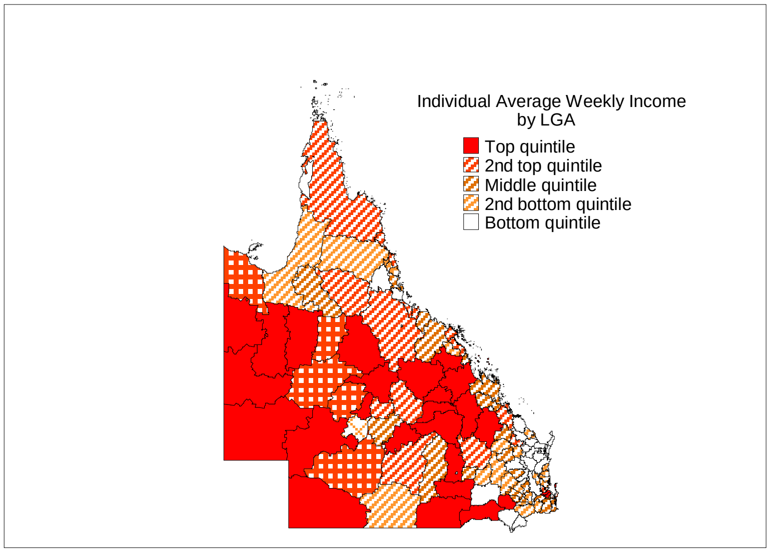

Figure 1. Average Income in Queensland, as at 2001 Population Census

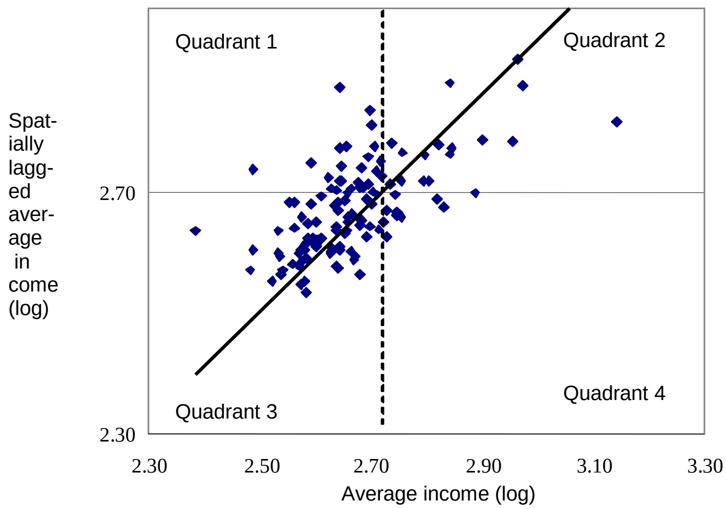

Figure 2. Moran Scatterplot- Regional Incomes

Table 2. Moran I Tests for Spatial Autocorrelation

Test	Moran *I*
Value	8.6224
Marginal probability	0.0000

Further insight into spatial clustering can be gained through the use of local indicators of spatial association. These latter measures include the Ord and Getis (1995) *G*-stat and the Local Moran *I* Statistic (see Anselin, 1995). In this section the Local Moran *I* statistic is used to analyse the spatial clustering. The Local Moran *I* statistic is an extension of the Moran *I* statistic, and decomposes the global measure into contributions for each location, referred to as LISAs or Local Indicators of Spatial Association.

This statistic can be represented as:

$$I_i = x_i \Sigma w_{ij} x_j \tag{1}$$

where x_j is the difference between the value of the variable under consideration for area *i* and the mean value for that variable, w_{ij} is a weight representing the strength of connection between areas *i* and *j*, developed from neighbour information. The Local Moran *I* statistic can be used to identify those "hot spots" where there is a significant spatial clustering of similar values of the variable. The values of I_i are positive when values at neighbouring locations are similar and negative if they are dissimilar.

The significance map presented in Figure 3 indicates regions with Local Moran *I* statistics that are significant at the 1 and 5 percent levels of significance. In this significance map it can be seen that the clustering of high levels of income in central Queensland (identified in the quintile map of Figure 1) is significant, as is the clustering of low income areas in the Wide Bay – Burnett statistical division, north of Brisbane.

The results of this analysis suggest that levels of regional income are not randomly allocated across geographic space. Instead, the quintile map, Moran scatter diagram and Moran *I* statistic all suggest that the variable is spatially dependent, while the local Moran *I* provides an indication of where significant clusterings of similar levels of income occur. Overall, this suggests that geographic location cannot be ignored, as in much empirical research conducted at the regional level, further, this may need to be considered in our formal model of regional income determination.

3. AN EMPIRICAL MODEL OF REGIONAL INCOME

This section develops a simple model of regional income variation. In this model, regional income is hypothesised to be a function of four basic characteristics, being; the demographic profile of the region (*D*), the educational profile of the regional population (*E*), the regional industry structure (*S*) and the general economic conditions of the region, in this case proxied by the regional unemployment rate (*U*). Thus the model can be specified as:

$$INCOME = \beta_0 + \beta_1 D + \beta_2 E + \beta_3 S + \beta_4 U + \varepsilon \tag{2}$$

In this equation the β are regression coefficients to be estimated and ε is an error term. Additionally, analysis conducted in the previous section has suggested that regional variation in income is not randomly allocated across geographic space, so that geographic location may also need to be considered in the model.

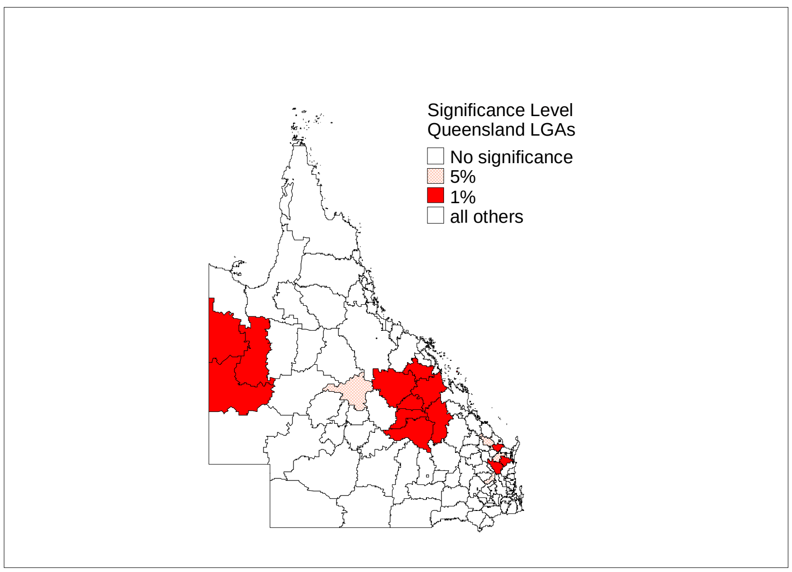

Figure 3. Significance Map of Average Income in Queensland LGAs

In this study, an estimate of regional income (INCOME) has been derived using a measure developed by Needleman (1978) with the log of regional income being used. Regional population may be an important control variable, regions with larger populations and labour forces, may have a deeper labour market, i.e. in addition to a greater diversity of industries, such regions may have proportionally more high level positions than would be expected in small regional economies. For example, small towns may have their bank managers but regional centres will have regional managers, etc. For this reason regional population may be an important determinant of regional income differences and so, for this reason our model incorporates the log of regional population (LPOP).

Other control variables, used in this study to capture the regional demographic profile, comprise the proportion of the labour force that are females (PERFEM), the proportion of the population that are indigenous persons (PERIND), mature aged persons (PER45+) and youth[3] (PERYNG). Typically, higher proportions of females are engaged in part-time employment and for this reason would be expected to have lower levels of income. For indigenous persons, data from the 2001 census indicates that in Queensland, indigenous persons are highly represented in low skilled occupations with 27.1 percent employed in the ASCO Labourers category, while overall only 7.9 percent of the Queensland labour force is in this category.

The two variables to capture differences in the age profile of regional labour forces were included because it is considered likely that age and skill level are closely related. PERYNG provides a measure of the number of 15 to 24 year olds in the regional labour force. Young persons are more frequently employed on a part-time basis or in lower skilled occupations. In contrast, mature aged workers tend to have accumulated skills over their working life, but this may be offset by lower rates of labour force participation or higher levels of part-time work than prime aged persons[4], for this reason the variable PER45+ being the proportion of the population aged over 45 has been included.

Within labour economics, it is generally believed that low skilled workers will have lower incomes than high skilled workers. For this reason the proportion of the population with bachelor and above level qualifications (PERBACH) and certificate level vocational training (PERCERT) have been included. If higher levels of education increase the productivity of workers this should be seen in their higher income levels. Furthermore, education has been found to be a significant determinant of regional economic growth (see, for example, Chatterji, 1998 or Lucas, 1998). In addition, education levels have been found to influence productivity (see for example, Black and Lynch, 1996 and Mason and van Ark, 1994) and differences in labour productivity are likely to result in differences in wages (see Weiss, 1995).

Differences in the proportion of regional employment in specific industries may also influence regional income levels, particularly if these industries have relatively different levels of income from that of the average. For this reason, the percentage of the labour force employed in Agriculture (PERAGR) and the percentage of the labour force employed in mining (PERMIN) have been included in this study. Agricultural incomes are subject to large fluctuations, and recent droughts and low world commodity prices would suggest that agricultural incomes will be below average. For this reason, regions with a high proportion of the labour force involved in agricultural industries may have lower average incomes. On the over hand, it is well known that mining incomes are well above the average, for this reason it would be expected that regions with a large

---

[3] The proportion of the population aged 15 to 24 or youth in the ABS definition, while those over 45 are considered mature aged.

[4] It is of course impossible to include the proportion of prime aged persons as this would result in perfect collinearity between these variables.

proportion of their labour force employed in this industry would have relatively high average incomes.

The final variable incorporated in this study of regional income is the regional unemployment rate (UERATE). This study has used ABS data from the 2001 Census of Population and Housing to derive average regional incomes. The data include incomes from all payments including unemployment benefits. For this reason the unemployment rate of the regional economy may be a significant determinant of cross regional variation in the level of income received within a region and it would be expected that regions with higher rates of unemployment will have lower incomes. In addition, some studies (see for example Partridge and Rickman, 1997) have found that high unemployment rates act to reduce regional wages with the excess labour supply acting as a disincentive for employers to offer higher wages.

4. ECONOMETRIC METHODOLOGY

Regional science has always recognised the role of space in determining regional economic performance. Space is also increasingly being recognised in empirical modelling through the use of techniques that formally incorporate a role for geographic location. These techniques allow the specification and testing of models that incorporate geographic spillover effects, or specify a dependence between observations at different points in geographic space. Collectively, these techniques form the field of spatial econometrics and range from simple descriptive statistics, which can be used to determine if similar values of a variable are clustered together in geographic space, through to methods for the estimation of structural equations that formally recognise the role of geographic location.

In this study two types of spatial econometric models were considered, being the spatial lag and error models.The spatial lag model takes the form:

$$\begin{aligned} &Y = \rho W y + X\beta + \varepsilon \qquad (3) \\ &\varepsilon \approx N(0, \sigma^2 I_n) \end{aligned}$$

while the spatial error model is defined as:

$$\begin{aligned} &Y = X\beta + \mu \\ &\mu = \lambda W \mu + \varepsilon \qquad (4) \\ &\varepsilon \approx N(0, \sigma^2 I_n) \end{aligned}$$

where Y is a vector of N observations of the dependent variable, X is a $N \times K$ matrix of observations of the explanatory variables, β is a vector of regression coefficients, ε is a vector of residuals, μ is an independently and normally distributed error term with constant variance, and W is an $N \times N$ spatial weight matrix. Anselin (2002) notes that these models require specialised estimation techniques, such as maximum likelihood or instrumental variables. In this study, maximum likelihood techniques, implemented through the R statistical software package have been used.

The weighting matrix W shows the interconnectedness of the areas in the sample; each element W_{ij} in W tells us the strength of interaction between the pair of regions *i* and j. Generally, it is expected that neighbouring areas would have a stronger interaction (larger W_{ij}) compared to geographically distant areas. Magalhaes *et al*. (2000) note that the main reason for the use of the spatial weight matrix is to associate a variable, at one point in geographic space, to the observation of the variable in other spatial locations. In contrast to time series, where the relation in time can be expressed by the simple notion of a lag operator L, where $L^s y_t = y_{ts}$ shifts y_t,s periods back in time, in space the problem becomes more complicated. The additional complication stems from the fact that there are many possible directions over which the spatial shift operator can be applied. One solution that has been offered to this problem is the use of the concept of a spatial lag operator L_s, with the idea being to use a weighted sum of the values of neighbouring units.

In this study a first order spatial weight matrix has been used. In this case, a symmetric matrix is defined by having the element (i, j) set equal to 1 if *i* and *j* are neighbours and *0* otherwise. By convention, the diagonal elements are set to zero, i.e. $w_i = 0$. Before use in estimation the weight matrix is standardised, denoted by the superscrip*t s*, with each of the non-zero elements being defined as $W_{ij}^s = w_{ij} / \Sigma_j w_{ij}$. In this matrix, the elements of the rows sum across to one. This manipulation facilitates the interpretation of the weights as an averaging of neighbouring values and also ensures the comparability between models of the spatial parameters in many spatial stochastic processes (Anselin and Bera, 1998).

The spatial lag model, shown in equation (3), is related to the distributed lag interpretation of time series economics. The lagged dependent variable, Wy, can be seen as equivalent to the sum of a power series of lagged dependent variables stepping out across a map, with the impact of spillovers declining with successively higher powers of ρ. This may be termed a structural autoregressive relationship, and one would expect it to be based on economic processes[5].

An alternate specification might be the spatial error model, shown in equation (4). This model presupposes a shared spatial process affecting all variables. This spatial process is frequently interpreted as indicating missing variables. In this model, λ is the residual spatial autocorrelation coefficient and represents unmodelled shocks. These sorts of effects include regional characteristics that are not part of the model but affect neighbouring regions similarly. Anselin (1999) notes that this type of regression incorporates a special case of a non-spherical error term. In this situation, OLS remains unbiased, but it is no longer efficient and the classical estimation of standard errors will be biased.

---

[5] For an explanation see Anselin (2002)

From this discussion it can be seen that the inclusion of spatial effects into an applied econometric model is typically motivated either on theoretical grounds, following the formal specification of spatial interaction in an economic model, or on practical grounds, due to the peculiarities of the data. In the empirical model represented by equation (2) no role has been assigned to spatial interaction in the model. For this reason, it is suggested that spatial effects may be important only in practical terms, due to the impact that their omission may have on the error term of the estimated equation and its impact on the bias or precision of the coefficient estimates.

5. MODEL ESTIMATION AND EVALUATION

The first step in modelling involved the estimation of the model using conventional OLS techniques that have frequently been used in empirical regional modelling incorporating cross sectional data. Specialised techniques have been developed to study geographically related data and the next stage of the investigation involved determining whether spatial autocorrelation was present in the residuals of the equation estimated using OLS, and if so, whether it is best represented by a spatial lag or spatial error model.

A series of tests have been developed and the results derived from the application of some of these tests are presented in Table 3. The tests used here comprise the Moran *I* statistic, Lagrange Multiplier (LM) error and LM lag tests and robust versions of these tests.

Table 3. Tests for Spatial Autocorrelation.

Test	Test stat	p-value
Moran I test	2.7814	0.0027
LM error test	4.6337	0.0314
LM lag test	20.9970	0.0000
Robust LM error test	0.0627	0.8023
Robust LM lag test	16.4260	0.0001

The results of the Moran *I* test suggest that we can reject the hypothesis of spatial independence due to the small marginal probability associated with this test and so conclude that the residuals from the OLS estimation exhibit spatial dependence. The Moran *I* test is perhaps the most commonly used specification test for spatial autocorrelation. Anselin *et al*. (1996) note that this test consistently outperforms other tests in terms of power in simulation results. A limitation of the test however, is that it provides no indication of whether the spatial autocorrelation present in the residuals is due to a true spatial process best represented by a spatial lag model, or an error process, best represented by a spatial error model. On the other hand, the LM tests provide a means of discriminating between the spatial lag or error model.

Table 4. OLS and Spatial Lag Model Estimates.

	OLS model			Spatial lag model		
	Estimate	t value	Pr(>|t|)	Estimate	z value	Pr(>|z|)
(Intercept)	2.8418	17.2780	0.0000	1.9669	8.8728	0.0000
LPOP	0.0028	0.2200	0.8260	0.0059	0.5307	0.5957
PERFEM	0.0723	0.3230	0.7473	-0.1795	-0.9155	0.3599
PERIND	-0.2018	-3.6830	0.0004	-0.1215	-2.4416	0.0146
PERYNG	-0.1813	-0.5920	0.5550	-0.1313	-0.4958	0.6200
PER45+	-0.7288	-5.3090	0.0000	-0.4782	-3.8246	0.0001
PERBACH	0.7439	2.7480	0.0070	0.8246	3.5169	0.0004
PERCERT	0.4111	1.9150	0.0580	0.6028	3.2095	0.0013
PERAGR	0.1791	2.9260	0.0042	0.1504	2.8356	0.0046
PERMIN	0.5804	8.1040	0.0000	0.4510	6.9430	0.0000
UERATE	-0.4503	-2.5440	0.0123	-0.5350	-3.4966	0.0005
ρ	n.a.	n.a.	n.a.	0.3061	5.1550	0.0000
	Residual standard error =0.04801			Log likelihood = 219.16		
	R^2 = 0.83 $\overline{R}^2$ = 0.81			R^2 = 0.85 $\overline{R}^2$ = 0.84		
	F (10, 114), 53.38, p value = 0.0000			F (11, 113), 60.54, p value = 0.0000		
				LM test for residual auto = 0.6008		
				p value = 0.4383		
	Goldfeldt - Quandt test = 1.1161			Goldfeldt - Quandt test = 1.1492		

The LM error and lag tests shown in Table 3 suggest that the spatial lag version of the model is the most appropriate with the LM lag and the robust version of this test both significant. Table 4 presents the estimated coefficients and diagnostics of the OLS version of the model and the spatial lag model. The second third and forth columns in this table provide the results from the model estimated using OLS, while columns 5, 6 and 7 provide the coefficients and z-values, of the maximum likelihood version of the spatial lag model.

The results for the OLS model and the spatial lag version of the model indicate that the variables explain over 80 percent of the cross regional variation in average income. Further, the *F* statistics for both of these estimated equations are significant at normal levels. In addition, it can be seen that the test for residual spatial autocorrelation of the error term of the spatial lag model suggests that the lag specification has overcome this problem. Further, the results from the application of the Goldfeldt - Quandt test to both the OLS and spatial lag model suggest that heteroscedasticity is not a problem with this model specification, with the critical value of this test being 1.84 in this situation.

The results presented in Table 4 indicate that certain variables incorporated in the model are not significant. LPOP (the log of regional Population), PERFEM, (percentage of females in the labour force) and PERYNG (percentage of 15-24 year old persons) were not significant in either the OLS or the spatial lag model. In contrast the remaining variables were significant and generally their coefficients had the expected sign. PERIND (the proportion of the regional population that were indigenous), PER45+ (percentage of the population over 45) and UERATE (the unemployment rate) have a negative sign, indicating that higher shares of the regional population of indigenous origin or mature aged, and higher rates of unemployment were associated with lower incomes.

The remaining variables, i.e. PERMIN (the proportion of the regional labour force employed in the mining industry), PERAGR (the proportion of the regional labour force employed in agriculture), PERCERT and PERBACH (the proportion of the regional population with a certificate or bachelor and above level qualification respectively), all had positive coefficients indicating that larger values of these variables were associated with higher regional incomes.

For the most part these results are not surprising, an exception to this being the significant positive coefficient of PERAGR which suggests that higher shares of employment in agricultural industries is associated with higher incomes. Queensland has experienced several years of drought and low commodity prices, for this reason it would be expected that higher shares of employment in agriculture would be associated with lower incomes. However, this is refuted by the census data which indicates that agricultural income is higher than the average income for Queensland, i.e. the average income of persons employed in agriculture was $598 in 2001 compared to the average income in Queensland at this time of $478. Hence it can be seen why higher shares of persons employed in agriculture are associated with higher regional incomes.

Also of significance was the sign and significance of the coefficients of the variables incorporated to capture differences in the education profile of the regions. Both variables, PERBACH and PERCERT (the proportion of the population with bachelor and above level qualifications and certificate level qualifications respectively) were found to be significant determinants of the cross regional variation of income. Both of these variables estimated coefficients were positive, indicating that higher proportions of the population with these types of qualifications are associated with higher levels of income.

The relative size of the coefficients is also of interest with the coefficient of PERBACH (0.82) indicating that a 1 percentage point increase in the population with a bachelor degree is associated with a 0.82 percent increase in the level of regional income. In contrast, regional income is not as sensitive to changes in the proportion of persons with a certificate level qualification with the coefficient of this variable being 0.60, indicating that a 1 percentage point increase in the proportion of the population with this level of qualification is associated with a 0.60 percent increase in the level of regional income. This result is not surprising considering that the average income of persons with bachelor level and above qualifications tends to be higher. However, the results also indicate that regional industry policy which includes occupations requiring trades and other certificate level qualifications will also positively address regional disparities in average income.

6. CONCLUSIONS

This paper has examined the determinants of variation in regional income across small areas of Queensland. Various explanatory variables were considered, including variables pertaining to the demographic profile, educational profile, regional industry structure and general economic conditions of each of the LGAs. Concerns of spatial dependence and heteroscedasticity made testing for them necessary, as they may have resulted in biased or inefficient estimators if inappropriate techniques were used.

Initial examination of the map of income quintiles amongst LGAs, as well as the Moran scatter plot hinted at spatial clustering. The Global Moran I test of income confirmed that spatial association was very significant. A Local version of the Global Moran I test identified which areas of Queensland appeared to have significant spatial dependency. This led to the conclusion that the geographical location could not be overlooked in the analysis and that the income of a LGA did depend on the income of neighbouring regions. This was also confirmed in the analysis of the residuals of the OLS version of the model where spatial autocorrelation was found to be significant. The tests suggested that the spatial lag version of the model was the most appropriate for the dataset and the results from this model suggested that it had accounted for this problem.

By and large the model outlined in section 2, was confirmed with most of the variables being significant and of the hypothesized sign. An exception to this was the sign of the coefficient of the variable PERAGR (the proportion of the labour force employed in agricultural industries), which was positive contrary to expectations. However, further investigation has revealed that agricultural income was substantially above the average income in Queensland at the time of the 2001 census.

Finally, it was found that both variables incorporated in the model to capture differences in the education profiles, or quality of the regional labour force, were significant and positive. This suggests that higher levels of both certificate level qualifications and bachelor and above level qualifications are associated with higher regional incomes. This result also suggests that regional industry policy which includes occupations requiring both higher education as well as trades and other certificate level qualifications will positively address regional differences in average income.

REFERENCES

Aigner, D.J., and Heins, A.J. (1967) On the determinants of income inequality. *American Economic Review*, 57, pp. 175-184.

Anselin, L. (1995) Local indicators of spatial association_LISA. *Geographical Analysis*, 27, pp. 93-115.

Anselin, L. (1999) *Spatial Econometrics.* Bruton Center, School of Social Sciences, University of Texas at Dallas, Richardson.

Anselin, L. (2002) *Under the hood, issues in the specification and interpretation of spatial regression models*. REAL Technical Paper Series, No. 02-T-09. Regional Economics Applications Laboratory, South Mathews, Urbana, Illinois.

Anselin, L. and Bera, A. (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah (ed), *Handbook of Applied Economic Statistics*. Giles: Marcel Dekker.

Anselin, L., Bera, A., Florax, R. and Yoon, M. (1996) Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), pp. 77-104.

Black, S.E., and Lynch, L.M. (1996) Human-Capital Investments and Productivity. *Technology, Human Capital, and the Wage Structure,* 86(2), pp. 263-267.

Chatterji, M (1998), Tertiary Education and Economic Growth. *Regional Studies,*.32, pp. 349-354.

Garofalo, G., and Fogarty, M.S. (1979) Urban income distribution and the urban heirarchy-equality hypothesis. *Review of Economics and Statistics*, 61(3), pp. 381-388.

Levernier, W., Rickman, D. and Partridge, M.D. (1995) Variation in U.S. state income inequality. *International Regional Science Review*, 18(3), pp. 355-378.

Long, J.E, Rasmussen, D.W., and Haworth, C.T. (1977) Income inequality and city size. *Review of Economics and Statistics*, 53(2), pp. 244-246.

Lucas, R.E. (1998) On the mechanics of economic development. *Journal of Monetary Economics,* 22, pp. 3-42.

Mason, G. and van Ark, B. (1994) Vocational Training and Productivity Performance: An Anglo-Dutch Comparison. *International Journal of Manpower, 15(5),* pp. 55-69.

Maxwell, P. and Peter, M. (1988) Income inequality in small regions: A study of Australian Statistical Divisions. *The Review of Regional Studies*, 18(1), pp. 19-27.

Magalhaes, A., Hewings, J.D. and Azzoni, C. (2000) *Spatial dependence and regional convergence in Brazil*. Regional Economics Application Laboratory, Technical paper series, No. 00-T-11, University of Illinois.

Messner, S. and Anselin, L. (2004) Spatial Analysis of Homicide with Areal Data. In M. Goodchild and D. Janelle (eds.), *Spatially Integrated Social Science*, Oxford University Press: New York. (also available at http://agec221.agecon.uiuc.edu/users/anselin/papers/smla.pdf).

Molho, I. (1992) Local Pay Determination. *Applied Economics,* 6(2), pp. 155-194.

Needleman, L. (1978) On the approximation of the Gini coefficient of concentration. *The Manchester School*, 46, pp. 105-122.

Nord, S. (1980) Income inequality and city size: An examination of alternative hypothesis for large and small cities. *Review of Economics and Statistics*, 62(4), pp. 502-508.

Ord, J. and Getis, A. (1995) Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, 27, pp. 286-306.

Partridge, M.D. and Rickman, D.S. (1997) The dispersion of U.S. State unemployment rates: the role of market and non-market factors. *Regional Studies*, 31(6), pp. 593-606.

Reza, A.M. (1974) Geographical differences in earnings and unemployment rates. *Review of Economic and Statistics,* 60, pp. 201-208.

Shah, A. and Walker, M. (1983) The distribution of regional earnings in the UK. *Applied Economics,* 15, pp. 507-519.

Topel, R.H. (1986) Local labour markets. *Journal of Political Economy*, 94(3), pp. S111 - S143.

Topel, R.H. (1994) Regional labour markets and the determinants of wage inequality. *The American Economic Review,* 8(2), pp.17-22.

Weiss, W. (1995) Human capital vs signaling explanation of wages. *Journal of Economics Perspectives*, 9(4), pp. 133-154.