Influence of urban form on transit behaviour in the Auckland region: A spatial Durbin analysis

Abstract

Based on cross-sectional data, this paper contributes to the existing literature by offering an insight into the spatial structure of the public transport sector in New Zealand. By decomposing the total effect of one explanatory variable into direct and indirect effects，the use of spatial Durbin model provides a better understanding of the urban form factors that influence bus mode share. The results show that the total effects comprised mostly of the spill-over impacts, and only a relatively small portion is attributed to the direct effects on bus mode share that arose from own-region changes in any given variable.

Table of Contents

1. Introduction

Throughout the world, as people's incomes rise, many shift to faster, more comfortable and more individually flexible means of transportation (Downs, 2003). Not surprisingly, like most modern cities, the commuting pattern in Auckland is dominated by the automobile, with almost 86% of the share for the morning journey to work (JTW) attributed to private motor vehicles, while public transport (PT) accounted for around 7% of the journeys in 2007. In comparison to other Australasian cities, a recent ranking confirms Auckland's weak position in terms of PT use, with only 46 PT trips per capita per annum, while Wellington generates almost twice this number at 91, and Sydney has almost threefold (ARTA, 2009 [1]; Statistics New Zealand, 2010). Auckland is thus inevitably characterised by an elevated level of car-dependence and an extremely low PT patronage.

With the aim of reducing automobile dependence and inducing non-automobile commuting, transport planners around the world are attempting to tackle the travel growth problem by implementing transport planning projects that can promote forms of sustainable urban development (e.g. Banister and Marshall, 2000; Barton *et al*., 1995). Without exception, transport authorities in Auckland have also implemented several major projects to facilitate the development of PT, from both smaller-scale initiatives such as expanding bus priority lanes to large-scale development such as bus and rail infrastructure projects. Therefore, from the perspective of local government and urban planners, it is crucial to have a solid understanding of how well the design and layout of urban areas do in terms of contributing to a reduction in automobile use and PT travel promotion.

The motivation behind this paper is that in order to properly understand the relationship between urban form and transit ridership, it is necessary to consider the associated spatial structures more specifically. Over the past few decades a number of studies have attempted to identify the impact of urban form on different travel behaviours such as mode choices, travel demand and travel patterns (e.g. Gordon *et al*., 1989; Headicar and Curtis, 1994; Kitamura *et al*, 1997; Næss *et al*., 1995). Unfortunately, all the above studies share one shortcoming in common. These analyses assume that observations are independent of one another in a geographical context. However in reality, it seems unlikely that region *i*'s transport network

[1] ARTA, Auckland Regional Transport Agency, has been superseded by Auckland Transport since the reorganisation at local government on 1st November, 2010. However, this paper still refers to ARTA since all of the data used here were compiled when ARTA was in existence.

in terms of vehicles and public transport infrastructure is independent of that of its neighbouring region *j*. Furthermore, from the econometric point of view, ignoring spatial characteristics between observations could, in turn, produce biased and inconsistent estimators (LeSage and Pace, 2010). This general limitation from past literature gives rise to the need for alternative spatial estimation approaches, such as the spatial Durbin model (SDM), as it has the advantage of separating total effect of a particular variable on the transit ridership into own-region and neighbourhood effects.

The remainder of the paper is organised as follows: section two provides a review of spatial effects and details the structure of spatial regression models. Section three describes the dataset, outlines the variables used, and specifies the regression models employed. Section four presents the empirical results and the final section provides a conclusion by summarising the key findings of the study.

2. Review of Spatial Effects and Spatial Durbin Model

2.1 Spatial dependence and spatial heterogeneity

Recently, the economics literature has paid extensive attention to spatial issues when conducting theoretical and applied econometric studies using cross-sectional data of a geographic nature. According to Anselin (1988a), spatial data are mainly characterised by two features: spatial dependence (or spatial autocorrelation) and spatial heterogeneity (or spatial non-stationarity). [2] Together, these two particular features have been regarded as spatial effects and nowadays they are perceived as major challenges in spatial analysis (Du and Mulley, 2006).

LeSage and Pace (2010) emphasise that data collected from nearby areas are commonly interdependent with each other, thus this spatial dependence requires special consideration when doing research because the consequence of ignoring this structure could result in biased estimates. As is well established in the literature, spatial dependence can exist in two forms, substantive and residual (Anselin, 1988a). The former is caused by spatial correlation of observed features, which indicates that the explanatory variable in one geographical space is correlated with the variable in adjacent or nearby geographical space. The latter, on the other

[2] This paper will use the synonyms, spatial dependence and spatial autocorrelation, and spatial heterogeneity and spatial non-stationarity, interchangeably.

hand, relates to the fact that spatial autocorrelation can also be found among unobserved variables when the error terms are correlated across contiguous geographical space (Case, 1991). As noted in Tobler's (1979) first law of geography, "near things" are more related than "distant things". As a result, spatial dependence appears to be the best known spatial effect and acknowledged most often in literature (Anselin, 1988a).

The other spatial effect, spatial heterogeneity, as described in Can (1990), refers to the systematic variation in the behaviour of a given process across space, and usually leads to heteroscedastic error terms. It could exist in a dataset which contains spatial information because unless a geographical space is uniform and boundless, every location will have some degree of uniqueness relative to the other locations (Getis *et al.*, 2004).

Practically, it remains difficult to fully disentangle the effects of spatial non-stationarity from spatial dependency (Bailey and Gatrell, 1995). Moreover, Anselin (2010) further refines this point by introducing the "inverse problem" concept, where spatial heterogeneity becomes particularly challenging since it is often difficult to separate it from spatial dependence. Florax and Nijkamp (2003) advocate that in fact, the occurrence of spatial heterogeneity does not necessarily have severe implications for the information that can be obtained from a spatial data series. Spatial dependency, on the other hand, however, does, because an observation is partly predictable from its neighbouring observations. This study only focuses on part of the spatial dependency effect.

LeSage (2004) advocates the use of spatial models when dealing with spatial effects. He argues that a conventional regression augmented with variables representing geographic dichotomous information, such as region dummies, or variables reflecting interaction with locational coordinates, which allow variation in the parameters over space, can hardly ever outperform a spatial model.

2.2 Spatial Durbin model

In practice, one should realise that spatial autocorrelation can have effects on both dependent and explanatory variables. Hence a "mixed" spatial Durbin model (SDM) introduced by Anselin (1988a) offers a more flexible alternative and might be more appropriate to apply by including the "inherent spatial autocorrelation" and "induced spatial dependence" simultaneously (Osland, 2010).

The SDM is specified as follows:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\gamma + \mathbf{u} \qquad \textbf{(1)}$$

The reduced form of equation (1) is:

$$\mathbf{y} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \rho\mathbf{W})^{-1}\,\mathbf{W}\mathbf{X}\gamma + (\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{u} \qquad \textbf{(2)}$$

In this case, an additional term $\mathbf{WX}\gamma$ must be included in the model to capture the $k \times 1$ autoregression coefficient vector γ of the spatially lagged explanatory variables **WX**, which measures the marginal impact of the explanatory variables from neighbouring observations on the dependent variable **y** (Kissling and Carl, 2008).

Furthermore, Osland (2010) argues that this SDM could be developed from either a spatial error model (SEM) (Anselin, 2006) or from a spatial autoregressive model (SAR) (Bivand, 1984), and this "mixed" model can be viewed as an unrestricted model of either SEM or SAR. In other words, the SDM further nests the SAR and the SEM by involving spatial dependence in the error term as well as in the dependent variable.

According to LeSage and Pace (2009), SDM is the only model that will produce unbiased estimates regardless of the true data-generation process (i.e. whether it is a spatial lag or a spatial error model). This is why the SDM is often viewed as the dominant spatial model among others. However, public transport studies incorporating spatial effects are relatively scarce compared to their rich applications in other fields, such as agricultural and resource economics (e.g. Benirschka and Binkley, 1994; Hurley *et al.*, 2001; Roe *et al.*, 200; Weiss, 1996) and housing and real estate analysis (e.g. Basu and Thibodeau, 1998; Berg, 2002; Case *et al.*, 2004; Pace and Gilley, 1997; Smith and Wu, 2009).

To the best of our knowledge, Greer and van Campen (2011) have produced the only published paper which specifically takes spatial effects into account when analysing the determinants of work trip bus ridership in the context of New Zealand using the SEM model.

It concludes that after adjusting for spatial dependency, the SEM model represents a significant improvement over the simple OLS model by providing more accurate estimates of parameter values and improves the predictive power of the model.

However, there remains a potential weakness in interpreting Greer and van Campen's results. In addition to the spatial lag of the dependent variable included on the right hand side of the regression equation, it seems plausible that neighbouring area units' characteristics, such as population density and rush hour frequency, could also play a significant role in explaining variations in a given area unit's bus ridership. This implies that further investigation of the impact of lagged explanatory variables on transit ridership is required. This study applies the SDM model which has the ability to capture the characteristics of neighbouring regions in order to account for any influence they may exert on their neighbour's transit ridership patterns.

3. Data and Empirical Models

3.1 Data

The major source of data for this study was the New Zealand Census, collected and complied by the Statistics New Zealand on the census day, 6th March, 2006. Additional data, such as distance to Auckland's CBD, distance to the nearest rail or ferry terminals and census area unit land areas, are calculated using ArcMap. Furthermore, the rush hour frequency, which combines the total number of buses running through and stopping within each area unit, during both morning and afternoon peak hours, is compiled using ArcMap and Microsoft Excel. The data were geocoded at the centroid of each area unit.

The census area unit is the second smallest geographical unit defined by Statistics New Zealand. Area units are aggregations of meshblocks, and they are non–administrative areas that are in between meshblocks and territorial authorities in size (Statistics New Zealand, n.d.). All data used in this study were compiled at this geographical level. In line with Yu *et al.* (2010), smaller units such as the meshblocks would render too much variation, and consequently increase analytical instability, while larger units such as territorial authorities would aggregate data too much and are thus incapable of providing useful results. [3] There are

[3] There are more than ten thousand mesh blocks and only seven territorial authorities in the Auckland region.

398 census area units within the Auckland region, of which sufficient data could be collected on 318. The final analysis is consistent with the dataset employed by Greer and van Campen (2011). [4]

3.2 Variables

The selection of variables is mainly inspired by previous bus patronage studies. The dependent variable Bus_i is the percentage of workers in area unit i who take bus as their main transport to work, self-reported on the census day. It was obtained by dividing the total number of bus passengers by the total number of JTW commuters in the i^{th} area unit. The percent mode share to bus offers an overall measure of the prominence of bus transport in the Auckland region.

Figure 1 presents the spatial distribution of bus mode share in the Auckland region based on 2006 census data. From this figure, it is evident that the bus mode share is not evenly distributed across area units. More specifically, the observations do not seem to be randomly distributed over space. Area units which have a high level of bus mode share, represented by the darker colour zones, tend to be closely concentrated in the centre, while the area units which have a relatively low bus ridership, shown in the lighter colour parts, are scattered around the boundaries.

Additionally, small clusterings of high values are also detected on the northeast and southeast corners of the map, which further indicates the spatially heterogeneous nature of the distribution of bus mode share. Therefore, spatial autocorrelation is apparently observed, because undoubtedly the probability of a specific value of the bus mode share variable in one specific location (area unit) depends on its value in neighbouring locations.

[4] Note that in Greer and van Campen (2011), the total number of observation is 318 area units. However, Waiheke Island is dropped from the dataset because although there are some bus services running within this area, they are not connected with any other bus services due to its isolated nature.

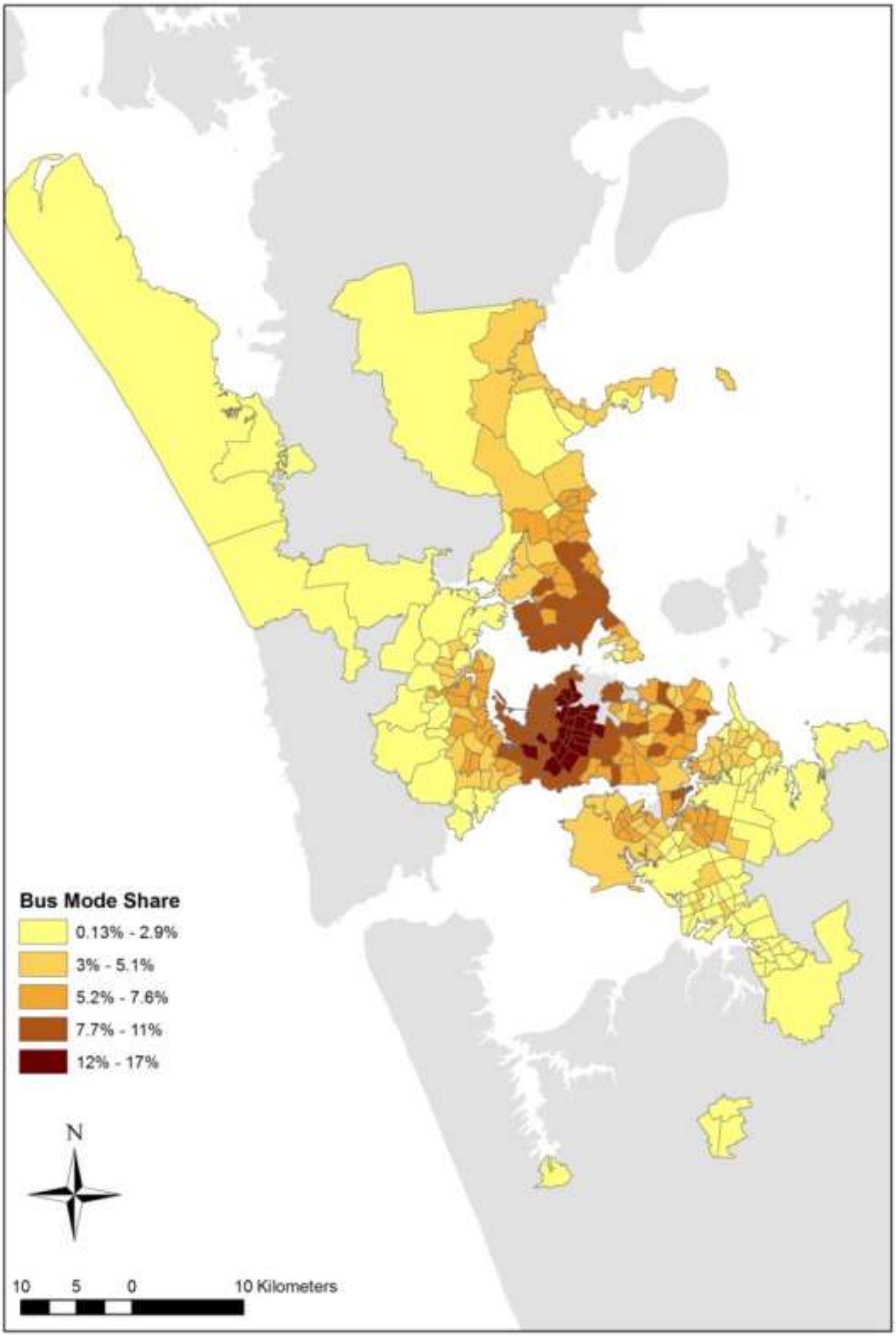

Fig. 1. Spatial distribution of bus mode share in the Auckland region

Potential bus mode share predictors are divided into three categories: urban form (UF), transit service (TS), and demographic and socioeconomic characteristics (DS). The final dataset includes eight independent variables, where:

1. UF variables:

- $PopD_i$: gross population density in the i^{th} area unit in the Auckland region, measured by the total number of inhabitants per square kilometre;
- $EmpC_i$: employment density, measured by the total number of full-time and part-time employees per capita in the i^{th} area unit in the Auckland region;
- $Dwelling_i$: total number of private owner occupied dwellings in the i^{th} area unit in the Auckland region; to be used as an indicator of land use patterns;
- CBD_i: distance to CBD from the centroid of the i^{th} area unit in the Auckland region, in kilometres;

2. TS variables:

- $Station_i$: distance to the nearest PT terminal/stop other than bus (either train or ferry) from the centroid of the i^{th} area unit in the Auckland region, measured in kilometres;
- $Freq_i$: frequency of bus service within the i^{th} area unit in the Auckland region;

3. DS variables:

- $Income_i$: median household income measured in thousands of New Zealand Dollars (NZD) within the i^{th} area unit in the Auckland region;
- Car_i: mean number of motor vehicles per household within the i^{th} area unit in the Auckland region;

The Transportation Research Board (1996) points out that urban form variables, such as road network type and neighbourhood type, along with variables such as in-vehicle time and an indicator of the waiting environment which describe the quality of transit service indicated by Paulley *et al*. (2006), also influence the demand for PT; unfortunately, these data are not available. A summary of key descriptive statistics of the variables used in this analysis are presented in Table 1. As can be seen from this table, the bus share for JTW trips in the Auckland region is fairly low; the average figure for all 317 area units is only 5.65%, ranging from a low of 0.13% to a high of 17.43%.

Table 1. Area unit level descriptive statistics of variables for Auckland Region

Variables	Mean	Std. Dev.	Minimum	Maximum
Dependent variable				
Bus (%)	5.65	3.44	0.13	17.43
***UF* explanatory variables**				
PopD (per km^2)	833.98	405.64	1.47	1726.74
EmpC (per capita)	0.48	0.08	0.27	0.66
Dwelling	1241.25	503.51	114	3270
CBD (km^2)	16.68	8.36	2.23	43.29
***TS* explanatory variables**				
Station (km^2)	3.67	4.17	0.14	35.53
Freq	130.03	94.48	2	476
***SD* explanatory variables**				
Income (in thousands of NZD)	27.11	6.32	14.4	48.4
Car	1.71	0.2	1.18	2.32

3.3 Empirical bus mode share models

A logarithmic transformation is applied to both dependent and explanatory variables with the intention of capturing the *a priori* belief that *ceteris paribus*, the impact of each explanatory variable on bus mode share is diminishing (Bresson *et al.*, 2004; Gomez-Ibanez, 1996).

Therefore, firstly, the non-spatial bus mode share model in log-log form is specified as below:

$$\mathbf{lnBus = X\beta_{\mathrm{OLS}} + \varepsilon_{\mathrm{OLS}}} \qquad \mathbf{(3)}$$

The above equation posits that the variation in the natural logarithm of the bus mode share (*lnBus*) in area unit *i* is explained by the variables in matrix **X**, which includes a constant term, the natural logarithm of UF variables (*PopD*, *EmpC*, *Dwelling* and *CBD*), the natural logarithm of TS variables (*Station* and *Freq*), and the natural logarithm of DS variables (*Income* and *Car*). Since the linear regression (8) is estimated by ordinary least squares, it is labelled as the OLS model and hence the estimated results serve as a benchmark against the following spatial model estimations. [5]

Secondly, the following SAR model is:

$$\mathbf{lnBus} = \rho\mathbf{WlnBus} + \mathbf{X\beta}_{\mathrm{SAR}} + \mathbf{u}_{\mathrm{SAR}} \qquad \textbf{(4)}$$

Similarly, the SEM is:

$$\mathbf{lnBus} = \mathbf{X\beta}_{SEM} + \boldsymbol{\varepsilon} \qquad \textbf{(5)}$$

$$\text{where } \boldsymbol{\varepsilon} = \theta\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u}_{\mathrm{SEM}}$$

Lastly, the SDM is given as:

$$\mathbf{lnBus} = \rho\mathbf{WlnBus} + \mathbf{X\beta} + \mathbf{WX}\gamma + \mathbf{u}_{\mathrm{SDM}} \qquad \textbf{(6)}$$

4. Estimation Results

4.1 Spatial weights matrix

In empirical spatial econometric models, the selection of a spatial weights matrix, normally denoted as **W**, plays an important role. LeSage (2002) outlines the many possible ways to quantify the structure of spatial dependence between observations. Typical approaches include: distance decay (Anselin, 1980), structure of a social network (Doreian, 1980), economic distance (Case *et al*., 1993) and *k* nearest neighbours (Pinkse and Slade, 1998). However, as Leenders (2002) illustrates, one major challenge facing spatial econometric models is that the spatial weights matrix **W** cannot be directly estimated but needs to be explicitly specified *a prior*, and current economic theory provides no formal guidance for this. Although a wide range of literature, echoed by Anselin (2002), has proposed several approaches to create the spatial weights matrix, there barely exists a formal guidance on how to select the "optimal" spatial weights as existing specifications all seem somewhat arbitrary.

[5] All of the models (i.e. OLS, SAR and SDM) are estimated using *Stata* 11.

Practically, in spite of their lesser theoretical appeal, geographically derived weights are among the most widely applied specification in spatial econometric analysis (Anselin, 2003a). In addition, as Manski (1993) argues, this popularity of geographically derived weights is due to the fact that the structure of **W** is constrained so that the weights are truly exogenous to the model, thus avoiding identification problems.

Generally there are two types of geographically derived weights based on proximities, namely, a binary measure of continuity (when two areas share common borders) and a continuous measure of distance. Following a majority of empirical studies (Fingleton, 1999, 2000; Le Gallo, 2002; Le Gallo *et al*., 2003; Rey and Boarnet, 2004), this paper uses a two-dimensional Cartesian coordinate system with the ordered pair (x, y) coordinates to create a spatial weights matrix **W** based on the distance decay specification along with its eigenvalues matrix **E**.

By convention, the weights matrix **W** has been row-standardised such that every row of the matrix sums to one (i.e. $\sum_j w_{ij} = 1$). Each element of **W** is therefore defined as:

$$w_{ij} = 0 \text{ if } i = j$$
$$w_{ij} = \frac{1}{d_{ij}} \text{ if } d_{ij} \le d^* \text{ and;}$$
$$w_{ij} = 0 \text{ if } d_{ij} > d^* \text{ if observation } i \ne j$$

where d_{ij} is the spherical distance between the centroids of area units i and j, and d^* is the critical cut-off distance. This inverse Euclidean distance, d_{ij}, contains a maximum threshold band of 24.14 kilometres to guarantee connections between all area units, that is, each spatial unit must have at least one neighbour. [6] This indicates that two area units are considered neighbours when the distance between their centroids is less than 24.14 kilometres, and not neighbours if their centroids lie 24.14 or more kilometres apart.

4.2 Moran's *I* test

A univariate Moran's *I* test for residuals is the most commonly employed first-step specification test for spatial autocorrelation (Moran, 1948; Anselin, 1999). The test does not

[6] The default unit for cut-off length is in miles in *Stata 11*, by conversion, 15 miles are approximately equal to 24.14 kilometres.

specify an explicit alternative spatial model (i.e. either SAR or SEM models) but has power against both (Anselin and Rey, 1991).

The Moran's *I* test for residuals in matrix notation is captured by:

$$\mathbf{I} = (N / \mathbf{S_0})(\mathbf{e}'\mathbf{W}\mathbf{e} / \mathbf{e}'\mathbf{e})$$

where $\mathbf{e}$ denotes a vector of OLS residuals, and $\mathbf{S_0} = \sum_{\mathrm{j}} \sum_{\mathrm{j}} w_{ij}$, a standardisation factor that corresponds to the sum of the weights for the non-zero cross products.

According to Florax and Nijkamp (2003), the interpretation of Moran's *I* should be parallel to a correlation coefficient; however the major distinction is that its value is not bounded by the (-1,+1) interval. A positive value signals positive spatial autocorrelation, measuring the occurrence of similar levels of a variable being found over contiguous or nearby spaces. By contrast, a negative value signals negative spatial autocorrelation, measuring the joint occurrence of high and low attribute values in adjoining locations.

The Moran's *I* statistic shows a positive value of 18.733 with a p-value that is lower than 0.0001. As expected, this result indicates that the null hypothesis of no spatial dependence should be rejected. Furthermore, the test statistic indicates that positive spatial autocorrelation exists, and in order to obtain unbiased and consistent estimators, spatial models should be adapted instead of the non-spatial OLS estimations.

4.3 The Lagrange Multiplier test

By applying the Lagrange Multiplier (*LM*) test, we select between a spatial lag and a spatial error alternative (Anselin, 2003a). Basically there are two major forms of the LM test. The LM_{lag} statistic tests the null hypothesis of no spatial autocorrelation in the dependent variable; the LM_{error} statistic, on the other hand, tests the null hypothesis of no significant spatial autocorrelation in the error terms.

The LM test against a spatial lag alternative (LM_{lag}) is demonstrated in Anselin (1988b) and takes the following form:

$$LM_{lag} = [\mathbf{e'Wy} \,/\, (\mathbf{e'e}/N)]^{2} / \mathbf{D}$$

where $\mathbf{D} = [(\mathbf{WX\,\beta})'(\mathbf{I}_n - \mathbf{X}(\mathbf{X'X})\text{-}(\mathbf{WX\,\beta})/\sigma^2)] + \mathrm{tr}(\mathbf{W}^2 + \mathbf{W'W})$. [7]

By contrast, the LM test against a spatial error alternative (LM_{error}), which is originally outlined in Burridge (1980), takes the form of:

$$LM_{error} = [\mathbf{e'We} \,/\, (\mathbf{e'e}/N)]^{2} / [\mathrm{tr}(\mathbf{W}^2 + \mathbf{W'W})]$$

Apart from a scaling factor, this statistic corresponds to the square of Moran's *I*.

As recommended by Florax and Nijkamp (2003), if both hypotheses can be rejected, one should consider constructing the robust forms of these LM tests which have the ability to correct for the presence of local misspecification of the other form (Anselin *et al*., 1996; Bera and Yoon, 1993). The test procedures of LM^r_{lag} and LM^r_{error} are identical to the one described above[8]. Both the classic and the robust LM tests are based on the residuals of the OLS model and are asymptotically distributed as $\chi^2(1)$.

Table 2 presents the diagnostics for spatial dependence. Under the classic LM test, both hypotheses of no spatially lagged dependent variable and of no spatially autocorrelated disturbances can be rejected at a 1% significance level. The robust LM tests consistently show the same results, with rejection of both hypotheses at a 1% significance level. This implies that OLS is rejected in favour of both SAR and SEM models. In addition, the statistic of LM^r_{lag}, 34.968, is greater compared to the result in LM^r_{error}, indicating a slight edge in favour of the spatial lag model.

Unlike what holds for the SAR's counterpart, the Autoregressive (AR) model in time-series analysis, the OLS estimation in the presence of spatial dependence will be inconsistent, simply because of the endogeneity issue discussed before. Therefore, in this study, the SAR and SEM models will be estimated using ML estimation (Ord, 1975; Anselin, 1988a).

---

[7] Where "tr" denotes the trace of the matrix *W*.

[8] The subscript "*r*" denotes "robust".

Table 2. OLS diagnostics for spatial dependence

Measure	Statistic	Significance
SEM		
LM_{error}	123.601	***
LM^{r}_{error}	30.395	***
SAR		
LM_{lag}	128.174	***
LM^{r}_{lag}	34.968	***

The results from the non-spatial OLS and the SAR model are reported in Table 3. Several distinctive points have been found. Firstly, consider the OLS result. [9] Overall, the coefficients of the urban form variables are significant and of the expected signs, in line with earlier findings in the literature. However, against expectations, the variable *logDwelling* is not significant.

The value of R-squared (R^2) is 0.730, indicating a reasonable model fit. [10] However, as the result from the Moran's *I* statistic and model diagnostic tests in Table 2 show, estimates using the OLS method suffer from a major problem: there is evidence of a positive spatial autocorrelation, and the *LM* test statistic suggests the lag specification as the appropriate alternative. Thus, the above OLS estimates should be interpreted with caution.

Firstly, in line with the conclusion from Moran's *I* statistic, the spatial autocorrelation coefficient estimate ρ for the SAR model is 0.762, and it is statistically significant at a 1% level, confirming the presence of positive spatial autocorrelation in the regression relationship. The OLS result simply ignores this spatial variation and produces biased estimates.

[9] The OLS result in Table 3 is not adjusted by heteroscedastic-robust standard errors so that they are compatible with the following spatial models.

[10] Adjusted R-squared for OLS estimation is 0.725.

Secondly, upward bias is found in most of the least-squares estimates, suggesting over-estimation of the sensitivity of bus mode share to the urban form, transit supply, and socio-economic and demographic characteristics when spatial dependence is disregarded. For instance, the OLS estimates imply that a one percentage increase in the population density in the area unit *i*, will increase the bus mode share in the same area unit by approximately 0.14%. However, after adjusting for spatial dependence, the estimated result from the SAR model suggests a much smaller effect of population density on the bus mode share, where a 1% rise in the population density in the area unit *i* will only increase the bus mode share in the same area unit by approximately 0.11%.

Thirdly, by taking the spatial lag into account, the fit of the model has improved dramatically. The R^2 statistic for the SAR model is 0.785, which has a higher value compared to the one in OLS. Therefore, after adjusting for spatial dependence, the overall fitness of the model has been improved.

Compared to the OLS, the signs and significance levels are maintained for all except the estimated coefficient on *log*(*EmpC*). Using the SAR model, the level of significance of this variable was reduced from 1% to 5%, and the impact of this variable on bus mode share also diminished by around one percent.

4.4 Choosing between alternative spatial dependence models

As Elhorst (2010) describes, if the OLS model is rejected in favour of both SAR and SEM models, then the SDM should be estimated. Therefore, a likelihood ratio (LR) test, also known as the score test, can subsequently be used to test two separate hypotheses that H_0: $\boldsymbol{\gamma} = 0$ and H_0: $\rho\boldsymbol{\beta} + \gamma = 0$.

Recall that the SDM model is reduced to the SAR model if $\boldsymbol{\gamma} = 0$. Osland (2010) proposes that when there is evidence of maintaining the SAR or SEM model, the SDM model specified by equation (5) and the following log likelihood tests may be useful in terms of determining the "true" spatial process. Thus, for the SAR model, one can determine the dominant model by testing the null hypothesis $\gamma = 0$. Rejecting the null hypothesis implies rejecting the SAR. Similarly, a common factor constraint: $\gamma = -\rho\boldsymbol{\beta}$ should be tested in order to determine the best model between the SDM and its SEM. Likewise, if the null is rejected, this indicates statistical evidence in favour of the SDM. With the aid of the LR test, one can decide the better model between the SDM and its restricted versions.

The likelihood ratio (λ) is defined as:

$$\lambda = 2[\ln (L_U) - \ln (L_R)] \sim \chi^2(\mathrm{m})$$

where L_U is the likelihood function of the unrestricted model (i.e. $L_U = L_{SDM}$) whereas L_R is the likelihood function of the restricted model (i.e. $L_R = L_{SAR}$ or L_{SEM}), and m is the number of restrictions imposed. The idea is that if the restrictions are valid, the log likelihood functions should appear to be similar in values and accordingly λ should be equal to zero.

The following results are obtained: $L_{SDM} = -59.929$, $L_{SAR} = -86.097$ and $L_{SEM} = -75.867$.

With 8 degrees of freedom, the critical values at 1%, 5% and 10% significance are 1.646, 2.733 and 3.490, respectively. The test statistics exceed the critical values for all cases, therefore we can reject the null hypothesis that the underlying spatial process is SAR or SEM at a 1% significance level; in other words, the restriction on parameter γ associated **WX** and also the common factor constraint are invalid. As a result, the unrestricted SDM should be employed to represent the data-generation process of the spatial autocorrelation. This result further implies that the spatial lags of both the dependent and explanatory variables should be included in the model. In fact, the inclusion of the spatial lags of explanatory variables makes reasonable sense as area units situated near each other should have similar values in terms of urban form, transit supply and socioeconomic/demographic variables, because economic activities tend to interact largely across space.

The estimation results for the SDM model are summarised in Table 3 below, alongside with the OLS and SAR estimates. Overall, the SDM explains over 85% of the variation in the bus mode share.

Table 3 Non-spatial OLS, spatial autoregressive model and spatial Durbin model (Dependent variable: lnBus)

Explanatory Variable	OLS Estimates		SAR Estimates		SDM Estimates	
Intercept	2.749	***	2.334	***	3.522	
log(PopD)	0.138	***	0.114	***	0.141	***
log(EmpC)	1.646	***	0.798	**	-0.285	
log(Dwelling)	-0.218		-0.132		-0.166	*
log(CBD)	-0.970	***	-0.440	***	-0.511	***
log(Station)	0.260	***	0.202	***	0.120	***
log(Freq)	0.158	***	0.118	***	0.143	***
log(Income)	-1.053	***	-0.610	***	-0.579	***
log(Car)	-0.865	***	-1.413	***	-0.732	***
Lag log(PopD)					-0.379	**
Lag log(EmpC)					3.528	
Lag log(Dwelling)					-0.066	
Lag log(CBD)					0.480	**
Lag log(Station)					0.098	
Lag log(Freq)					0.281	*
Lag log(Income)					0.410	
Lag log(Car)					-3.613	***
ρ			0.762	***	0.823	***
Squared Correlation	0.730		0.785		0.852	
Variance Ratio			0.833		0.779	
Log likelihood			-86.097		-59.929	

*** Estimated coefficients significant at 1% level; ** significant at 5%; * significant at 10%

4.5 Decomposing total effect into direct and indirect effects

Interpretation of the SDM model differs from that of its non-spatial regression counterpart, the ordinary least squares, as the k^{th} parameter vector $\boldsymbol{\beta}$ is no longer a partial derivative of $\mathbf{y}$ with respect to change in the k^{th} independent variable from the $n \times k$ matrix of $\mathbf{X}$ (LeSage and Fischer, 2008). Essentially, the spatial dependence components in the SDM expand the information set to include information from neighbouring area units. To see the impact of this, consider the partial derivative of the SDM in equation (7) with respect to a particular explanatory variable x_k:

$$\underbrace{\mathbf{M}}_{} = \underbrace{\frac{\partial y}{\partial x_k}}_{n \times n} = \underbrace{(\mathbf{I} - \rho\mathbf{W})^{-1}}_{n \times n}\underbrace{[\boldsymbol{\beta}_k + \mathbf{W}\gamma_k]}_{n \times n} \quad \textbf{(12)}$$

The partial derivative results in an $n \times n$ matrix **M** representing marginal effects, which is shown in equation (12). The impact on the dependent variable from a change in a coefficient can be decomposed into three ways, namely, direct, indirect and total effects. LeSage and Pace (2009) define the direct effect as the average of the diagonal elements of matrix M; it provides a summary measure that represents an average of the impacts on bus mode share arising from own-region changes in variable x_k. The indirect effect is defined as the average of the off-diagonal elements of matrix **M**; this effect is also known as the spatial spill-over effect as it measures the impact on bus mode share in area unit i arising from changes in variable x_k from all other area units. The total effect is calculated as the average row sums of matrix **M**; it includes both direct plus indirect effect. The total effect measures the average cumulative impact on each observation from changing the k^{th} explanatory variable by one unit across all observations.

Average direct, indirect and total effects estimated are reported in Table 4, along with inferential statistics (i.e. the figures in parenthesis are bootstrapped standard errors) calculated using a bootstrap method with 1,000 draws. Because all of the variables are expressed in natural logs, the coefficient estimates can be interpreted as elasticities.

Table 4. Direct, indirect and total effects of the spatial Durbin model

Variables	Direct Effect		Indirect Effect		Total Effect	
log(PopD)	0.136	***	-1.480	***	-1.345	***
	(4.73E-05)		(4.72E-05)		(5.81E-08)	
log(EmpC)	-0.230	***	18.543	***	18.322	***

	(5.30E-04)		(5.77E-04)		(1.05E-07)	
log(Dwelling)	-0.170	***	-1.141	***	-1.311	***
	(3.53E-05)		(3.54E-05)		(3.49E-08)	
log(CBD)	-0.510	***	0.335	***	-0.175	***
	(1.04E-05)		(1.03E-05)		(1.34E-08)	
log(Station)	0.124	***	1.108	***	1.232	***
	(3.43E-05)		(3.26E-05)		(5.10E-08)	
log(Freq)	0.151	***	2.245	***	2.395	***
	(7.03E-05)		(7.05E-05)		(1.16E-07)	
log(Income)	-0.580	***	-0.374	***	-0.955	***
	(1.13E-05)		(1.15E-05)		(7.01E-09)	
log(Car)	-0.816	***	-23.732	***	-24.548	***
	(7.51E-04)		(6.93E-04)		(8.37E-07)	

*** Estimated coefficients significant at 1% level

For the total effects, all estimated parameter values have the expected signs, with one exception for *log*(*PopD*). The total effects of *log*(*EmpC*), *log*(*Station*) and *log*(*Freq*) on transit ridership are all positive and significant; while the total effects of *log*(*Dwelling*), *log*(*CBD*), *log*(*Income*) and *log*(*Car*) and *log*(*PopD*) are negative and significant. Separating the total effect of a regressor into direct and indirect effects yields further insights.

Firstly, the positive sign on *log*(*EmpC*) suggests that as employment density from all sampled areas rises, the transit ridership will tend to fall. This outcome is comparable with the results in Paulley *et al*. (2006), where the density variables in this aforementioned work also respond positively to public transport demand. Secondly, for the two transit service variables, first of all, the total effect, which comprise the direct and indirect effects of *log*(*Station*), is positive and significant, implying that across the Auckland region, as the distance to train station and/or ferry terminal increases, commuters will prefer to choose buses as their transport mode. Next, both the direct and indirect effects of rush hour frequency show a significant positive effect on the bus mode share in a given area. This result provides insights to transport planning *viz*. that by increasing the number of buses during morning and peak hours, the effect will not only be reflected through a rise in the percentage of commuters who choose to take bus to work in its own district, but also an additional spill-over benefit which can be reflected in its nearby areas. The elasticity of total effect of this variable is about 2.4, which indicates that increasing the transit frequency in area unit *i* by one percent, the average bus mode share across all area units will rise by 2.4%, holding other variables constant.

Secondly, for the group of variables that respond negatively to bus mode shares, the parameter estimate on *log*(*Dwelling*) indicates that the larger the share of private owner occupied dwellings within an area unit, the lower the share of commuters who take bus, which seems intuitively plausible. The estimated coefficient on the total effect of the distance to the CBD is negative and significant, suggesting that the propensity to take bus decreases as the area unit is farther away from the CBD in the Auckland region. For the two demographic and socioeconomic variables, both the direct and indirect effects of income level exert a significant negative impact on the bus mode share, reflecting the idea that bus transport is an inferior good: as the commuters become wealthier, they will make fewer bus patronages for their JTW trips. Moreover, the direct effect of *log*(*Car*) show that there is an inverse relationship between the number of cars owned per households and the bus usage rate, which is in line with the findings on car ownership variable found in Zhao *et al*. (2006) and Vance and Hedel (2007). The indirect effect of cars exhibits the same tendency, suggesting that with a one percent rise in the number of cars owned in adjacent area unit *j*, the average bus mode share in any given area unit *i* tends to decline by approximately 23.7%. Therefore it is clear that the more private vehicles owned in a geographical confined region and around its neighbourhood, the less likely the commuters will chose to take bus to work, because they have a more convenient substitute.

Thirdly, the estimated coefficient on the total effect of *log*(*PopD*) is negative and significant at the 1% level. This implies that the population density from all observed area units affects negatively the percentage of workers who take bus as their main transport to work, which runs counter to our original hypothesis that high population density leads to high transit ridership.

As discussed earlier, total effect can be unravelled into direct (own-region) and indirect (spatial spill-over) effect. Some notable findings were revealed by our results: three urban form variables: *log*(*PopD*), *log*(*EmpC*) and *log*(*CBD*) have the opposite signs for direct and indirect effect parameters; while the rest stays consistent. The estimated coefficient on the direct effect of *log*(*PopD*) is positive and significant at the 1% level. The result is consistent with the assumptions made by previous studies without consideration of spatial effects (Maat *et al*., 2005; Steiner, 1994), where people living in high-density sectors prefer to use more public transport or walk more frequently, but will make fewer and shorter trips by private vehicles. However, the indirect effect is negative and also significant, suggesting that once the population density in nearby regions increases, the bus mode share in area unit *i* will decline. This outcome may be due to the fact that commuters in area unit *i* interpret the rise in population density in their neighbouring regions as a sign of potential congestion issues and/or dissatisfaction of the transit service, since buses might not be running on time, in such cases, taking private vehicles will be a better alternative than using public transport. Because the indirect effect is larger in magnitude, the total effect of *log*(*PopD*) is negative.

For the next urban form variable *log*(*EmpC*), the own-region effect of employment density exerts negative impact on the transit ridership, while the spatial spill-over effect is positive and significant, suggesting that if there are more employment opportunities in nearby regions, the bus mode share in region *i* will tend to rise. The negative parameter estimation of the direct effect of employment density reflects the greater attractiveness of low density suburban employment for the transit-dependent workers in the Auckland region. The positive indirect impact indicates that due to the spill-over effect commuters may find riding buses is a better option for longer trips, especially when workplace is far from the commuters' residential address. The sign of the total effect for this variable is negative because the magnitude of the negative direct effect outweighs the positive indirect effect.

Although the total effect of *log*(*CBD*) is negative, the direct and indirect effects have opposite signs. The direct effect of *log*(*CBD*) is negative and significant at the 1% level, suggesting that commuters are less willing to take the bus to work if they live farther away from Auckland city centre. Surprisingly, the estimated coefficient on the indirect effect of the distance to the CBD is positive, suggesting that the neighbourhood effect on the propensity to take the bus increases as the area unit is farther away from the CBD in the Auckland region. Although several possibilities were examined, there was no solution where this effect came out undesirable sign, therefore requires further investigation.

Another significant finding from the SDM output is that except for *log*(*CBD*) and *log*(*Income*), the total effects comprised mostly of the spatial spill-over impacts, and only a relatively small portion is attributed to the direct effects on bus mode share that arose from own-region changes in variable x_k. For instance, the indirect effect of *log*(*Car*) constitutes nearly 97% of the total impact of number of cars on bus mode share. Therefore, for the case of spatial dependence considered in the SDM model, least-squares regressions that ignore this spatial spill-over effect and only produce the coefficients that representing the summary impact measures, result in biased and inconsistent estimates. The result also reveals that spatial spill-overs dominate in transit behaviour analysis and greater attention should be paid from transport and urban planners on neighbourhood effect.

5. Conclusion

This paper estimated how urban form variables are related to bus mode share and how these effects vary across the Auckland region's diverse and dissimilar landscapes. Overall, based on area unit data, the analysis highlighted the complexity and importance of the spatial structure in determining the factors that influence the bus mode share.

The OLS method used in many transport-related studies assumes that the observations/regions are independent of one another in a geographical context. Thus, OLS looks for similarities in different spatial areas and effectively concentrates than in an 'average' figure to cover the whole space. However, this is not plausible when using spatially-defined data because it is likely to exhibit positive spatial autocorrelation, that is, correlation of a variable with itself through space. Ignoring the spatial characteristics between observations/regions will, in turn, produce biased and inconsistent estimators.

By conducting an in-depth case study using the Auckland region's data, urban forms, coupled with other factors that affect the bus mode share are explored and all these in turn are related under a spatial context. The Moran's *I* test shows that there is statistically significant evidence of the presence of positive spatial autocorrelation. Therefore, by taking into account the spatial dependence, the spatial regression models are selected over the non-spatial OLS model in order to obtain unbiased and consistent estimators. The empirical results show that the bus mode share in one area unit exhibits a positive relationship with the share in neighbouring area units.

However, the interpretation of these findings based on SAR and/or SEM models is confounded by the strong spatial autocorrelation of the urban form and other transit characteristics such as transit supply and socioeconomic/demographic differences across area units. By applying the likelihood ratio tests, this paper confirms the existence of spatial autocorrelation in the lags of both dependent and independent variables. This dominating spatial issue has been addressed by the use of the spatial Durbin model. Estimated results from SDM show that the total effects comprised mostly of spatial spill-over impacts, and only a relatively small percentage is attributed to the direct effects on bus mode share that arose from own-region changes in any given explanatory variable. For planners and developers, the SDM model is not only technically superior, but also preferable for evaluating policies and making investment decisions, as unlike traditional estimated coefficient interpretations, one can easily unravel total effect into own-region and spatial spill-over effect and. The results presented indicate that knowledge of a specific spatial lag may provide clues about the importance of future land use patterns on transit ridership.

References

Anselin L (1980) Estimation methods for spatial autoregressive structures. Regional Science Dissertation & Monograph Series No.8. Program in urban and regional science, Cornell University

Anselin L (1988a) *Spatial econometrics: Methods and models*. Kluwer, Dordrecht

Anselin L (1988b) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* 20: 1–17

Anselin L (1992) Spatial data analysis with GIS: An introduction to application in the social sciences. Technical Report 92-10, National Center for Geographic Information and Analysis University of California

Anselin L (1999) Interactive techniques and explanatory spatial data analysis. In: Longley PA, Goodchild MF, Maguire DJ, Rhind, DW (eds) *Geographical information systems: Principles, techniques, management and applications*, pp 251-264. John Wiley & Sons, New York

Anselin L (2002) Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27: 247-267

Anselin L (2003a) Spatial econometrics. In: Baltagi BH (ed) *A companion to theoretical econometrics*, pp 310-330. Basil Blackwell, Oxford

Anselin L (2003b) Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review* 26: 153-166

Anselin L (2006) Spatial econometrics. In: Mills TC, Patterson K (eds) *Econometric theory*, pp 901-969. Palgrave Macmillan, New York

Anselin L (2010) Thirty years of spatial econometrics. *Regional Science* 89: 3-25

Anselin L and Rey S (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23: 112–131

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

Auckland Regional Transport Authority (2009) Auckland's transport challenges. 2009/10 – 2011/12 Auckland regional land transport programme draft, Auckland Regional Transport Authority, Auckland

Bailey TC, Gatrell AC (1995) *Interactive spatial data analysis*. Longman, Harlow

Banister D, Marshall S (2000) *Encouraging transport alternatives: Good practice in reducing travel.* The Stationery Office, London

Barton H, Davies G, Guise R (1995) *Sustainable settlements*: *A guide for planners, designers and developers.* Severnside Research and Consultancy Unit, University of the West of England

Basu S, Thibodeau TG (1998). Analysis of spatial autocorrelation in housing prices. *Journal of Real Estate Finance and Economics* 17: 61-85

Benirschka M, Binkley JK (1994) Land price volatility in a geographically dispersed market. *American Journal of Agricultural Economics* 76: 185–195

Bera A, Yoon MJ (1993) Specification testing with locally misspecified alternatives. *Econometric Theory* 9: 649-658

Berg L (2002) Prices and constant quality price indexes for multi-dwelling and commercial buildings in Sweden. Working Paper 2002:2, Department of Economics, Uppsala University

Bivand R (1984) Regression modelling with spatial dependence: An application of some class selection and estimation methods. *Geographical Analysis* 16: 25-37

Bresson GJ, Dargay J, Madre, Pirotte A (2004) Economic and structural determinants of the demand for public transport: An analysis on a panel of French urban areas using shrinkage estimators. *Transportation Research Part A* 38: 269-285

Burridge P (1980) On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B (Methodological)* 42: 107–108

Can A (1990) The measurement of neighbourhood dynamics in urban house prices. *Economic Geography* 66: 254-272

Case A (1991) Spatial patterns in household demand. *Econometrica* 59: 953-965

Case A, Rosen HS, Hines JR (1993) Budget spillovers and fiscal policy interdependence: Evidence from the States. *Journal of Public Economics* 52: 285–307

Case B, Clapp J, Dubin R, Rodriguez M (2004) Modeling spatial and temporal house price patterns: A comparison of four models. *Journal of Real Estate Finance Economics* 29: 167-191

Doreian P (1980) Linear models with spatially distributed data, spatial disturbances or spatial effects. *Sociological Methods and Research* 9: 29–60

Downs A. (2003) *Still stuck in traffic: Coping with peak-hour traffic congestion*. Brookings Institution Press, Washington, DC

Du H, Mulley C (2006) Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. *Transportation Research Record* 1977: 197-205

Elhorst JP (2010) Applies spatial econometrics: Raising the bar. *Spatial Economic Analysis* 5: 10-28

Fingleton B (1999) Estimates of time to economic convergence: An analysis of regions of the European Union. *International Regional Science Review* 22: 5-34

Fingleton B. (2000) Spatial econometrics economic geography, dynamics and equilibrium: A third way? *Environment and planning A* 32: 1481-1498

Florax RJ, Nijkamp P (2003) Misspecification in linear spatial regression models. Discussion Papers No. 2003-081/3, Tinbergen Institute

Getis A, Mur J, Zoller HG (eds) (2004). *Spatial econometrics and spatial statistics*. Palgrave Macmillan, New York

Gomez-Ibanez J (1996) Big city transit ridership, deficits and politics: Avoiding reality in Boston. *Journal of the American Planning Association* 62: 30-50

Gordon P, Kumar A, Richardson HW (1989) Congestion, changing metropolitan structure and city size in the United States. *International Regional Science Review* 12: 45-56

Greer M, van Campen B (2011). Influences on public transport utilization: The case of Auckland. *Journal of Public Transportation* 14: 51-68

Headicar P, Curtis C (1994*)* Residential development and car-based travel: Does location make a difference? Proceedings of seminar C: Environmental issues, pp 117-130. 22nd PTRC European Transport Forum, Warwick

Hurley T, Kilian B, Malzer G, Dikici H (2001) The value of information for variable rate nitrogen applications: A comparison of soil test, topographical and remote sensing information. Technical report, Department of Applied Economics, University of Minnesota

Kissling WD, Carl G (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17: 59–71

Kitamura R, Mokhtarian P, Laidet L (1997) A micro-analysis of land use and travel in five neighbourhoods in the San Francisco Bay area. *Transportation* 24: 125-158

Le Gallo J (2002) Économétrie spatiale: L'autocorrélation spatiale dans les modèles de régression linéaire. *Économie et Prévision* 155: 139-157

Le Gallo J, Ertur C, Baumont C (2003) A spatial econometric analysis of convergence across European regions, 1980-1995. In: Fingleton B (ed) *European regional growth*, pp 99-129. Springer, Berlin

Leenders R (2002) Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks* 24: 21-47

LeSage JP (2002) Spatial econometrics. PhD topic lectures, Department of Economics, University of Toledo

LeSage JP (2004) Maximum likelihood estimation of spatial regression models. Lecture 1, Department of Economics, University of Toledo

Lesage JP, Fischer MM (2008) Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis* 3: 275-340

LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. Chapman & Hall/CRC, London

LeSage JP, Pace RK (2010) Spatial econometric models. In Fischer MM, Getis A (eds) *Handbook of applied spatial analysis: Software tools, methods and applications*, pp 355-376. Springer, Berlin

Maat K, van Wee B, Stead D (2005) Land use and travel behaviour: Expected effects from the perspective of utility theory and activity based theories. *Environment and Planning B: Planning and Design* 32: 33–46

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60: 531–542

Moran PAP (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society B (Methodological)* 10: 243-251

Næss P, Røe P G, Larsen S (1995) Travelling distances, modal split and transportation energy in thirty residential areas in Oslo. *Journal of Enviromental Planning and Management* 38: 349-370

Ord JK (1975) Estimation methods for models of spatial interaction. *Journal of American Statistical Association* 70: 120-126

Osland L (2010) An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Society* 32: 289-320

Pace RK, Gilley OW (1997) Using the spatial configuration of the data to improve estimation. *Journal of Real Estate Finance Economics* 14: 333-340

Paulley N, Balcombe R, Mackett R, Titheridge H, Preston J, Wardman M, Shires J, White P (2006) The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy* 13: 295-306

Pinkse J, Slade ME (1998) Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85: 125–154

Rey SJ, Boarnet MG (2004) A taxonomy of spatial econometric models for simultaneous equations systems. In: Anselin L, Florax RJ, Rey SJ (eds) *Advances in spatial econometrics: Methodology, tools and applications*, pp 99-119. Springer, New York

Roe B, Irwin EG, Sharp JS (2002) Pigs in space: Modeling the spatial structure of hog production in traditional and nontraditional production regions. *American Journal of Agricultural Economics* 84: 259–278

Smith TE, Wu P (2009) A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems* 11: 333-355

Statistics New Zealand. (2010). Subnational population estimates at 30 June 2010 (boundaries at 1 November 2010). Statistics New Zealand. Auckland

Statistics New Zealand. (n.d.) *Area unit*. Retrieved from: http://www.stats.govt.nz/surveys_and_methods/methods/classifications-andstandards/classification-related-stats-standards/territorial-authority/definition.aspx

Steiner RL (1994) Residential density and travel patterns: Review of the literature. *Transportation Research Record* 1466: 37–43

Tobler WR (1979) Cellular geography. In: Gale S, Olsson G (eds), *Philosophy in geography*. Reidel, Dordrecht

Transportation Research Board (1996) Transit and urban form. TCRP report 16, National Academy Press, Washington DC

Vance C, Hedel R (2007) The impact of urban form on car travel: Disentangling causation from correlation. *Transportation* 34: 575-588

Wall MM (2004) A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* 121: 311–324

Weiss MD (1996) Precision farming and spatial econometric analysis: Research challenges and opportunities. *American Journal of Agricultural Economics* 78: 1275–1280

Yu D, Peterson NA, Sheffer MA, Reid RJ, Schnieder JE (2010) Tobacco outlet density and demographics: Analysing the relationships with a spatial regression approach. *Public Health* 124: 412-416

Zhao F, Chow LF, Li MT, Liu X (2006) Transit ridership model based on geographically weighted regression. *Transportation Research Record: Journal of the Transportation Research Board* 1972: 105-114