Australasian Transport Research Forum 2013 Proceedings
2 - 4 October 2013, Brisbane, Australia
Publication website: http://www.patrec.org/atrf.aspx

Forecasting public transport demand for the Sydney Greater Metropolitan Area: a comparison of univariate and multivariate methods

Chi-Hong (Patrick) Tsai*

Institute of Transport and Logistics Studies, Business School, The University of Sydney

Email: chi-hong.tsai@sydney.edu.au*

**Corresponding author*

Corinne Mulley

Institute of Transport and Logistics Studies, Business School, The University of Sydney

Email: corinne.mulley@sydney.edu.au

Geoffrey Clifton

Institute of Transport and Logistics Studies, Business School, The University of Sydney

Email: geoffrey.clifton@sydney.edu.au

Abstract

Public transport demand forecasting is important for urban public transport planning. In the Sydney Greater Metropolitan Area (SGMA), bus and train as the two major public transport systems account for around two million trips per day. Understanding future changes in public transport demand in response to different policy scenarios gives important information for transport policy formulation. This paper reports forecasts of public transport demand in the SGMA using an Autoregressive Integrated Moving Average (ARIMA) model and a dynamic Partial Adjustment Model (PAM). The ARIMA model is estimated using monthly train and bus boarding data from 2007 to 2011. The PAM model estimates demand elasticities with respect to each of a number of public transport determinants, including the public transport fare, the socio-demographics of public transport users, the level of public transport service and land use characteristics. The PAM model is estimated using a pseudo panel dataset constructed from the Sydney Household Travel Survey from 1997 to 2009. The forecast accuracy of the two methods are compared to the actual demand observed in 2010 and 2011, using a holdout sample. The PAM model is then used for forecasting future public transport demand for the SGMA, for a number of policy scenarios. The forecasting results suggest that the ARIMA model can achieve better prediction accuracy in the short term, whereas a PAM model is preferred if the objective is to forecast future demand in response to various policy scenarios.

1. Introduction

Forecasting public transport demand is important because this is closely associated with urban transport planning and policy formulation. The methods of public transport demand forecasting are well developed in the literature and widely applied in transport planning models. However, there is little discussion on the relative merits of the various methods for public transport demand forecasting.

The conventional demand forecasting methods are generally categorised into univariate time-series approaches and multivariate demand modelling approaches, where the latter can be undertaken using a conventional four-step travel planning model or direct demand models. For public transport demand forecasting, the direct demand modelling approach has received more attention, given its capability of identifying the demand elasticity which represents the causal relationship between demand and explanatory variables. However, some forecasting studies in other domains of transport, such as tourist demand and fuel prices, have shown that the univariate approach demonstrates better prediction accuracy (du Preez and Witt, 2003; Li et al., 2010). The application of univariate time-series models are less evident in public transport demand forecasting, and thus the superiority of this forecasting method has not been demonstrated in this field.

The direct public transport demand models, which are widely employed in rail demand forecasting (Owen and Philips, 1987; Preston, 1991; Wardman and Tyler, 2000; Wardman, 2006; Blainey and Preston, 2010; Dargay et al., 2010), also require some update in the model specifications. There is an increasing number of studies highlighting the importance of a temporal effect of demand and the integration of public transport demand and land use characteristics. These relatively newly recognised elements of public transport demand determinants have not yet been commonly incorporated in previous forecasting models.

This paper applies both the univariate time-series method and the multivariate demand modelling method to forecast public transport demand in the Sydney Greater Metropolitan Area (SGMA). Section 2 reviews the literature on the previous forecasting studies and provides a methodological discussion. Section 3 conducts a univariate forecasting model based on the historical trend of demand changes. Section 4 presents a multivariate demand forecasting model incorporating the temporal effect and land use variables, with forecasting results being compared to the univariate model. Section 5 concludes the findings of this paper.

2. Literature review

The methods of travel demand forecasting can be generally categorised into univariate modelling and multivariate modelling approaches. The univariate modelling approach usually uses time-series models based on historical data to forecast future demand. This method assumes that patterns of the past demand will continue into the future with no exogenous determinants incorporated. Typically Autoregressive Integrated Moving Average (ARIMA) models are employed and this time-series modelling approach has been widely applied in transport studies, such as in rail freight volume (Babcock et al., 1999; Hunta, 2003), tourism demand (Burger et al., 2001; Lim and McAleer, 2002; du Preez and Witt, 2003), and energy demand (Ediger and Akar, 2007; Li et al., 2010). However, the univariate method is not commonly employed for public transport demand forecasting, although some studies have demonstrated that univariate time-series models can achieve better prediction accuracy than multivariate models (du Preez and Witt, 2003; Li et al., 2010).

Most studies on public transport demand forecasting usually adopt the multivariate modelling approach, in which the demand is predicted by a vector of explanatory variables and model inputs. Two principal forecasting methods in the multivariate approach are the multi-stage model and the direct demand model (Department for Transport2012). The former refers to the traditional four-step travel demand model consisting of trip generation, trip distribution, mode choice, and traffic assignment models, which has the advantage of being able to provide a comprehensive travel demand model across all transport modes and capture the effect of intervention on travel demand. For example, the Sydney Strategic Travel Model (BTS2011b) is a strategic travel planning model which is able to predict future demand in response to various policy scenarios such as land use changes and new public transport demand supply. However, the shortcoming of the multi-stage travel model is its limited capability of incorporating fine-grained land use characteristics and neighbourhood-scaled land use initiatives (Cervero, 2006). More practically, the multi-stage travel demand model requires comprehensive and detailed data which raises the cost of implementation and thus is not commonly used for forecasting minority modes such as rail and bus (Department for Transport2012), although this is not the case in Sydney.

The majority of studies on forecasting public transport demand use the direct demand modelling approach which aims to provide a causal statistical relationship between travel demand and its explanatory variables such as fares and quality of service. The key advantage of the direct demand models is the identification of demand elasticity, which represents how demand would be influenced by changes in the demand determinants. This approach has been extensively applied in rail demand forecasting (Owen and Philips, 1987; Preston, 1991; Wardman and Tyler, 2000; Wardman, 2006; Blainey and Preston, 2010; Dargay et al., 2010). Most of these studies investigate the relationship between rail patronage and rail fares, income, and the socio-demographics of travellers, as provide a number of demand elasticites with respect to each of the determinants. In addition, some studies have highlighted the importance of distinguishing short-run demand and long-run demand (Owen and Philips, 1987; Voith, 1991; Dargay and Hanly, 2002; Graham et al., 2009; Dargay et al., 2010; Kennedy, 2013; Tsai and Mulley, 2013 Forthcoming). These identify that travellers' behaviour in response to transport systems changes may not be immediate but instead take times to adjust, which is known as the temporal effect of travel demand.

Another growing body of literature on demand forecasting draws attention to the incorporation of land use variables and accessibility to public transport. Wardman and Tyler (2000) included an accessibility variable measured by distance to the local rail station in their rail demand model. The result shows that the demand elasticity with respect to accessibility is around -0.47 for leisure trips and -0.53 for business trips in the UK, suggesting its strong impact on rail travel demand. Cervero (2006) used the direct demand modelling to investigate the influence of Transit-Oriented-Development (TOD) characteristics on public transport usage in major cities of United States. These TOD characteristics, including land use density, feeder bus services and park and ride facilities, are measured at a neighbourhood level and are found to be significant in public transport demand. The literature on the connections between public transport demand and land use is substantial (Kitamura, 1990; Cervero and Kockelman, 1997; Cervero, 2002; Rodriguez and Joo, 2004; Tsai and Mulley, 2013 Forthcoming) confirming that public transport demand is significantly influenced by land use 3D (density, diversity, design) and accessibility although absent in most demand forecasting models identified in the literature.

3. Univariate time-series model

This paper uses a univariate model and a multivariate direct demand model to forecast public transport in the SGMA. An ARIMA model is employed for the univariate time-series analysis and is discussed in this section, and the dynamic Partial Adjustment Model (PAM), based on a pseudo panel dataset is constructed from the Sydney Household Travel Survey (SHTS), for the direct demand modelling approach incorporating land use 3D and accessibility measures is discussed in the next section, Section 4.

3.1 ARIMA model

A number of time-series models have been developed for forecasting. The literature does not specify an absolutely superior time-series model since the forecasting power depends on the nature of data and the context of study (Lim and McAleer, 2002). Among the time-series studies reviewed above, the ARIMA model, proposed by Box and Jenkins (1970), is the most popular time-series model given its capability of processing non-stationary and seasonal data. Instead of comparing the performance of the various univariate time-series models, this paper focuses on the forecasting power between the univariate modelling method and the multivariate modelling method and their implications for practical use of demand forecasting. Hence, the ARIMA model is selected for the univariate analysis of this paper given its popularity in demand forecasting studies and flexibility in a wide range of applications.

The ARIMA model, typically denoted as an ARIMA (p, q) model, consists of an autoregressive (AR) term, and a moving average (MA) term. The AR (p) model uses p lags of time to predict the dependent variable y as specified in equation (1).

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t \tag{1}$$

where t is the total number of time periods and e is the error term (white noise), $\emptyset_p$ is the parameter of the autoregressive variable at time period $t - p$.

The MA component uses q lags of error terms to smooth the time-series data and thus improves the forecast results. The MA (q) model is specified in equation (2), and the ARIMA model combining the AR and MA models is defined in equation (3)

$$y_t = e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_{t-q} \tag{2}$$

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_{t-q} \tag{3}$$

The ARIMA model in equation (3) is able to process non-seasonal data. If the data demonstrate seasonality, observed frequently in regular time series data, the seasonal effect will need to be controlled by employing a Seasonal ARIMA model. The general form of a seasonal ARIMA$(p, d, q)(P, D, Q)s$ model is specified as equation (4), where d is the order of non-seasonal differencing. P is the order of the seasonal autoregressive term. D is the order of the seasonal differencing. Q is the order of seasonal moving average process. s is the number of seaosonal cycles (i.e. s=4 more quarterly data; s=12 for monthly data).

$$(1-B)^d(1-B^s)^D y_t = \mu + \frac{\theta(B)\theta_s(B^s)}{\phi(B)\phi_s(B^s)} e_t \tag{4}$$

where

$$\phi_s(B^s) = 1 - \phi_{s,1}B^s - \ldots - \phi_{s,P}B^P$$

$$\theta_s(B^s) = 1 - \theta_{s,1}B^s - \ldots - \theta_{s,Q}B^Q$$

B is the seasonal difference operator; $(1\text{-}B^s)y_t = y_t - y_{t-s}$

μ is the constant

3.2 Data

Public transport demand in the ARIMA model of this paper is defined as the number of train and bus boardings per month. Monthly train and bus boarding data from 2007 to 2011 are provided by the Bureau of Transport Statistics (BTS) of Transport for New South Wales (TfNSW). The bus boarding data prior to 2007 are not available due to the institutional reform of the bus operating companies in the SGMA. Train and bus form the majority of public transport trips in the SGMA, and account for two million trips on an average weekday in 2011, including school buses. Other modes of public transport in the SGMA including ferries, light rail and monorail are excluded from this analysis because they account for less than 2.5 percent of total trips in the SGMA collectively.

Historical train and bus boarding statistics are shown in Figure 1. This identifies how there has been no dramatic demand change since 2007, although a strong seasonal effect can be observed. The historical trends of bus and train patronage are very similar, although train demand has been around 5 million trips higher than bus demand. The train demand has increased from around 20 million trips in 2007 to around 25 million trips per month. Bus demand has also increased by around 5 million trips since 2007 although from a lower base and both rail and bus trends show a slight addictive trend. This exploratory analysis indicates that the seasonal effect needs to be controlled in the ARIMA model and the forecast demand using the ARIMA model is expected to demonstrate a slightly increasing trend.

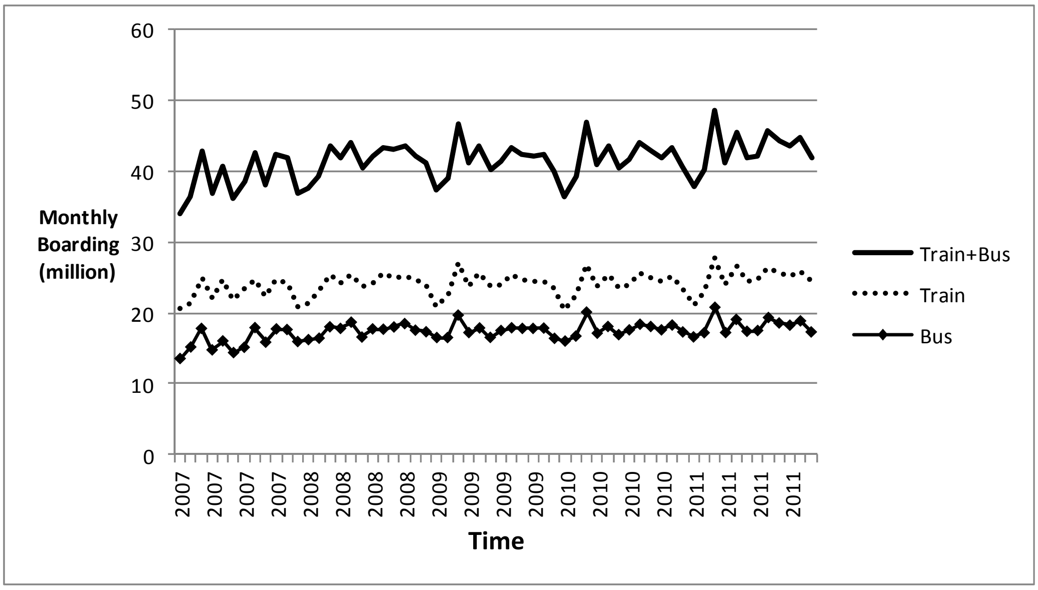
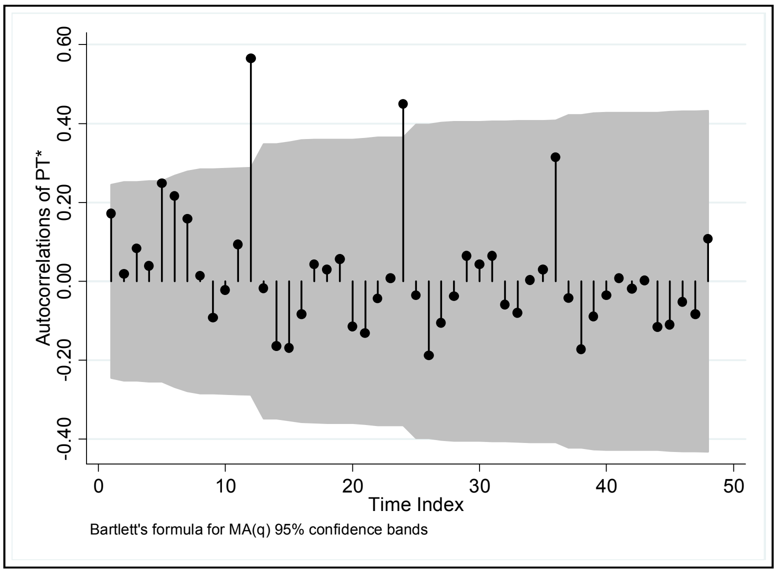


Figure 1. Monthly train and bus boarding statistics in the SGMA

3.3 Model identification

Time-series models assume that the data are stationary over time, that is, the mean, variance, and covariance of the data do not change over time. The stationarity can be observed from the autocorrelation function of the time-series data (Chatfield, 1989, page 18). Figure 2 shows the autocorrelation of monthly public transport demand since 2007. It can be clearly identified that the autocorrelation is significant at the 12th time lag and 24th time lag, which implies that the public transport demand in month t is significantly correlated with the demand in month $t-12$ and $t-24$. This strong seasonal effect needs to be adjusted through seasonal differencing to remove the non-stationarity as shown in equation (5).

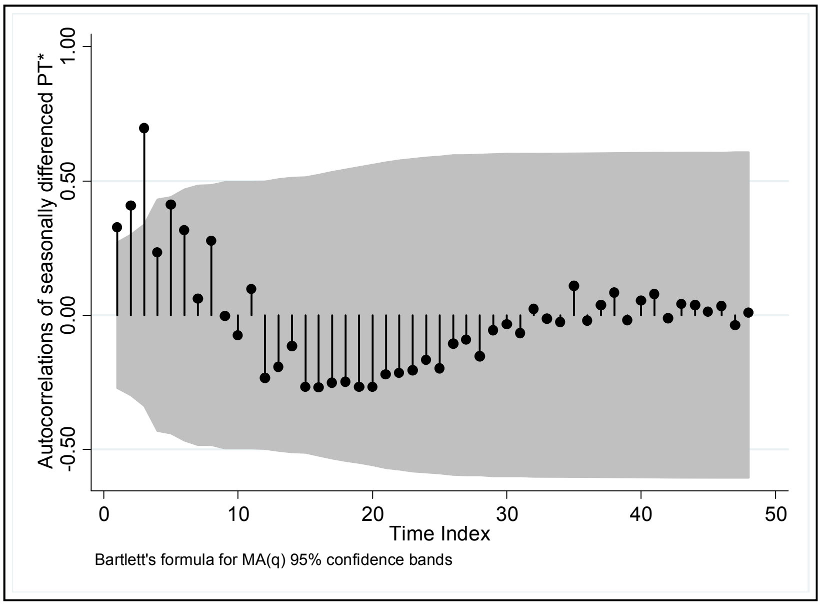
$$(1-B^{12})y_t = y_t - y_{t-12} \tag{5}$$



*PT: monthly public transport demand

Figure 2. Autocorrelation of monthly public transport demand

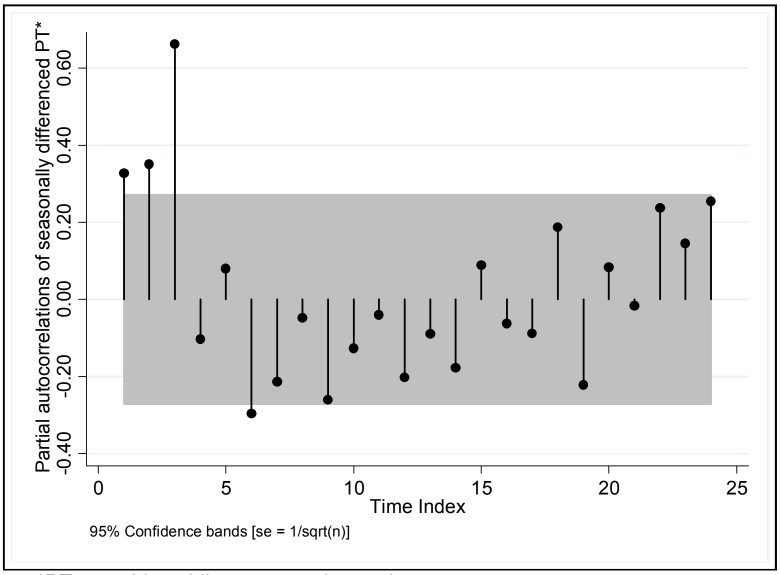
The autocorrelation of the seasonally differenced public transport demand in Figure 3 shows that the seasonal effect is removed after the seasonal differencing, and the autocorrelation drops off to insignificance after the third time lag which suggests that the time-series becomes stationary (Chatfield, 1989, page 20).



*PT: monthly public transport demand

Figure 3. Autocorrelation of seasonally differenced public transport demand

The orders of lags for the AR term (p) and the MA term (q) of the ARIMA model can be identified by examining the autocorrelation (Figure 3) and partial autocorrelation plots (Figure 4). As a general rule, the order of the AR term is determined by the number of partial autocorrelations that are significantly different from zero, and the order of the MA term is determined by the number of significant autocorrelations (Makridakis and Wheelwright, 1989, page 136). Figure 3 and Figure 4 each have three significant lags which indicate that the ARIMA model should include both of the AR and the MA terms.



*PT: monthly public transport demand

Figure 4. Partial autocorrelation of seasonally differenced public transport demand

The seasonal AR term (P) and MA term (Q) are usually determined through a try-and-error process, given that there are not sufficient autocorrelations and partial autocorrelations for precise identification as the non-seasonal terms (Makridakis and Wheelwright, 1989, page 139). Hence, various ARIMA models are evaluated in Table 1 and Model 13 is selected as the preferred model given its lowest Mean Squared Errors (MSE).

Table 1 Mean Squared Errors of ARIMA Models

Model	ARIMA$(p,d,q)(P,D,Q)s$	MSE[1]
Model 1	ARIMA (0,0,0)(0,1,0)12	3.32
Model 2	ARIMA (0,1,0)(0,1,0)12	4.17
Model 3	ARIMA (0,1,0)(1,0,0)12	6.42
Model 4	ARIMA(0,1,1)(2,0,0)12	3.69
Model 5	ARIMA (0,1,1)(3,0,0)12	3.52
Model 6	ARIMA (0,1,1)(0,1,1)12	2.37
Model 7	ARIMA(0,1,1)(1,0,0)12	5.22
Model 8	ARIMA(0,1,1)(2,0,0)12	4.85
Model 9	ARIMA (1,0,0)(0,1,0)12	2.98
Model 10	ARIMA (2,0,0)(0,1,0)12	2.71
Model 11	ARIMA (3,0,0)(0,1,0)12	1.78
Model 12	ARIMA (3,0,1)(0,1,0)12	1.72
Model 13	**ARIMA (3,0,1)(0,1,0)12**	**1.70**
Model 14	ARIMA (3,0,3)(0,1,0)12	1.79

[1]Mean Squared Errors

Model 13 is an ARIMA (3,0,1)(0,1,0)12 model with three lags of non-seasonal AR term and one lag of non-seasonal MA term, with a first lag of seasonal differencing. The model is diagnosed by checking the residuals and undertaking a Ljung-Box test (Ljung and Box, 1978), where the null hypothesis is that the residuals are independently distributed. The autocorrelations of the residuals in Figure 5 shows that only one residual is significantly different from zero which suggests that the residuals overall are not autocorrelated. The Ljung-Box test result suggests a p-value of 0.28 which fails to reject the null hypothesis. The model diagnostics confirm the randomness of residuals of Model 13.

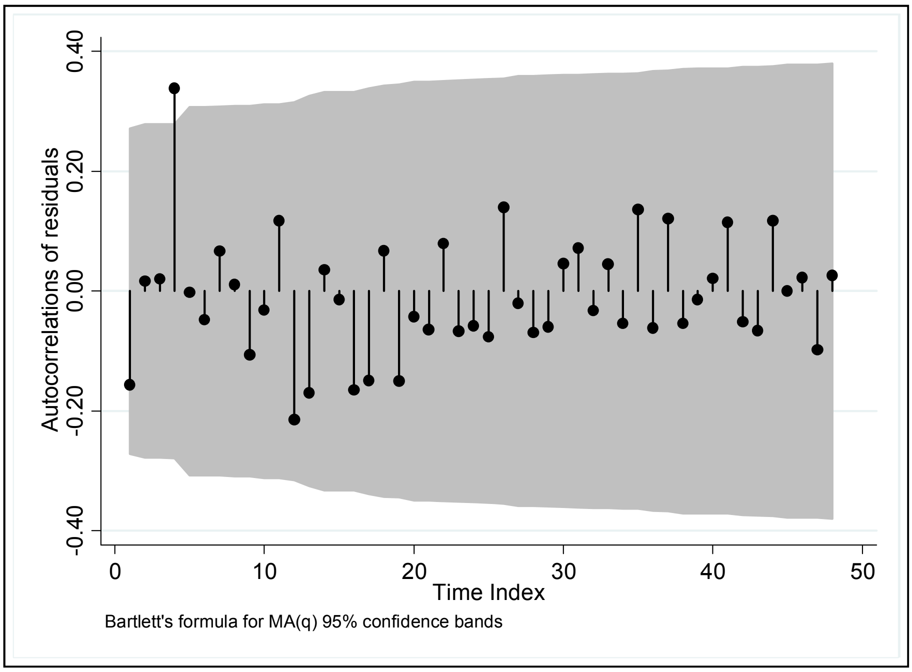


Figure 5. Autocorrelations of Residuals for Model 13

3.4 Forecasting results

Using Model 13 as the preferred model, the future public transport demand is forecast using a holdout sampling approach. Monthly data from 2007 to 2010 are used to estimate the ARIMA model, and the actual monthly public transport demand in 2011 is used as the base to compare to the forecast demand by the model. The monthly actual demand and forecast demand in 2011 are summarised in Table 2, which shows that the prediction difference varies between -2.48 percent and 1.79 percent on a monthly basis. The aggregate annual actual public transport demand in 2011 is 517.47 million trips and the forecast demand is 514.02 million trips, with an overall annual difference of -0.67 percent. The seasonal effect of the public transport demand is also predicted by the ARIMA model. The public transport demand in the SGMA is higher in the school seasons of March to May and August to November, except for April where the Easter Holidays reduces public transport use. This seasonal fluctuation is predicted by the forecasting model as shown in Figure 6.

Table 2. A Comparison of Actual Demand and Forecast Demand in 2011

Year	Month	Actual Demand (million)	Forecast Demand (million)	Difference
2011	Jan	37.81	37.29	-1.39%
2011	Feb	40.15	40.32	0.43%
2011	Mar	48.68	47.95	-1.50%
2011	Apr	41.17	41.90	1.79%
2011	May	45.59	44.46	-2.48%
2011	Jun	41.93	41.53	-0.95%
2011	Jul	42.10	42.56	1.09%
2011	Aug	45.61	45.12	-1.08%
2011	Sep	44.30	43.87	-0.96%
2011	Oct	43.51	42.99	-1.21%
2011	Nov	44.79	44.31	-1.07%
2011	Dec	41.84	41.72	-0.29%
2011	**Total**	**517.47**	**514.02**	**-0.67%**

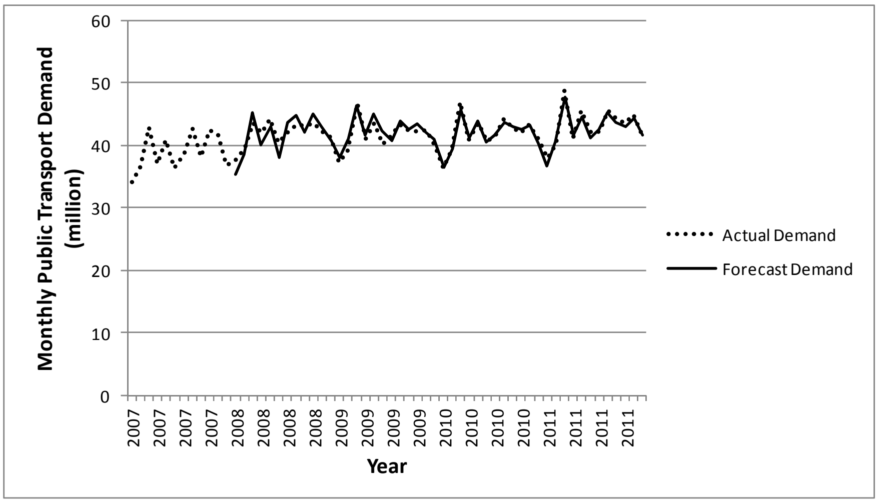


Figure 6. Actual Demand and Forecast Demand between 2007 and 2011

Based on the same ARIMA model, the future public transport demand between 2011 and 2026 are forecast as shown in Table 3 and Figure7. Assuming the pattern of demand shown in the historical trend between 2007 and 2011, the forecast for public transport demand in 2026 is 722 million trips with a total growth of 41.10 percent as compared to 2011. Figure 7 shows an addictive trend with seasonal effects similar to the historical trend, confirming the additive trend and seasonal effect are properly identified by the forecasting model in the long-term demand forecasting.

Table 3. Forecast Annual Public Transport Demand between 2011 and 2026

Year	Forecast Demand (million)	Growth Every 5 Year	Total Growth since 2011
2011	512.30	n/a	n/a
2016	580.40	13.29%	13.29%
2021	651.61	12.27%	27.19%
2026	722.86	10.94%	41.10%

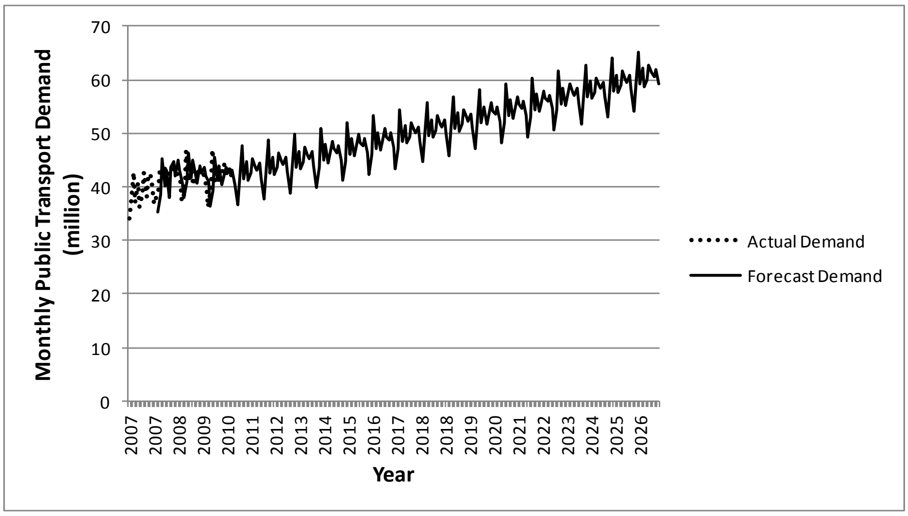


Figure 7. Forecast Public Transport Demand between 2007 and 2026

Despite the promising predictive power of the ARIMA model for 2011, as presented above, it must be remembered that the univariate forecasting model only replicates into the future the patterns which have been observed in the past. There is no causality employed between public transport demand and its explanatory factors and it may not be realistic to assume all the exogenous factors influencing current public transport will remain the same in the long-term future. Hence, the next section presents a multivariate direct demand model that creates a statistical relationship between public transport demand and its explanatory variables.

4. Multivariate Partial Adjustment Model

4.1 Pseudo panel data model

This section describes the multivariate direct demand model which is constructed to forecast public transport demand in the SGMA. The literature specifies that the determinants of public transport demand should include public transport trip price, socio-demographics, and quality of service (Balcombe et al., 2004). In addition, as the literature has suggested that the temporal effect and land use variables at a disaggregate geographical level should be taken into account in a public transport demand model, this paper employs a dynamic Partial Adjustment Model (PAM) using a pseudo panel dataset

constructed from the Sydney Household Travel Survey (SHTS) data between 1997 and 2010.
The pseudo panel data approach was proposed by Deaton (1985). Constructing a pseudo panel dataset involves assigning the respondents to groups by time-invariant variables, where the grouping criteria should create sufficient inter-group heterogeneity. Each created group consists of cohorts matching the grouping criteria over the observed time period. This approach has not been evident in the literature of public transport demand forecasting, but it has been increasingly applied in travel demand studies (Dargay and Vythoulkas, 1999; Dargay, 2001; Dargay, 2007; Weis and Axhausen, 2009; Tsai and Mulley, 2013 Forthcoming). As compared to conventional direct rail demand models where static models were employed (Preston, 1991; Wardman and Tyler, 2000; Wardman, 2006; Blainey and Preston, 2010), the pseudo panel data approach has the advantage of incorporating longitudinal data using household the travel survey data, and hence short-run and long-run demand elasticities can be estimated for a specific study area.

In this paper, the pseudo panel dataset are constructed from the SHTS data based on the birth year and the household distance to the Central Business District (CBD) as two grouping criteria. The created pseudo panel dataset consists of 20 groups over 13 years, with a final sample size of 256 cohorts after removing four cohorts with an average age of less than 18 years old, who were considered to have limited choices of trip modes and thus were excluded from the sample. The detailed construction process and the evaluation of the grouping criteria are comprehensively discussed in Tsai and Mulley (Tsai and Mulley, 2013 Forthcoming).

The PAM for predicting public transport demand is specified in equation (6):

$$\bar{D}_{g,t} = \bar{\beta}_0 + \lambda\bar{D}_{g,t-1} + \beta_P\bar{P}_{g,t} + \beta_E\bar{E}^{'}_{g,t} + \beta_S\bar{S}_{g,t} + \beta_L\bar{L}^{'}_{g,t} + \bar{u}_{g,t}\ ,\ \bar{u}_{g,t} = \bar{\alpha}_{g,t} + \bar{\varepsilon}_{g,t} \quad (6)$$

where the public transport demand $\bar{D}_{g,t}$ for a constructed group g in period t is determined by the public transport demand in period $t-1$ ($\bar{D}_{g,t}$) to capture the lagged demand adjustment, as well as public transport trip price $\bar{P}_{g,t}$, a vector of travellers' socio-economic factors $\bar{E}^{'}_{g,t}$, the level of public transport service $\bar{S}_{g,t}$ measured by service frequency, and a vector of land use characteristics and accessibility measures $\bar{L}^{'}_{g,t}$. $\bar{\beta}_0$ is the constant and $\bar{u}_{i,t}$ is the combined error term constituted of the unobserved group effect $\bar{\alpha}_{g,t}$ and the independent error term $\bar{\varepsilon}_{g,t}$. All variables represent the way in which the observation of each variable is the mean value for all individuals classified into group g in period t. A summary of the definitions and descriptive statistics of the variables is presented in Table 4.

Table 4. A Summary of Variables in the Public Transport Demand Model

Variable	Description	Unit	Mean	S.D.	Min	Max
Dependent variable						
PTTRIP	No. of bus and train trips per person per day	Trips/person	0.45	0.28	0.08	1.63
Price variable						
PRICE	Real public transport trip price	Dollars (AUD)	1.73	0.59	0.39	2.88
Socio-economic factors						
INCOME	Real annual personal income	Thousand dollars (AUD)	28.64	12.98	2.08	58.38
AGE	Age	Years	41.32	17.64	18.00	75.65
Public Transport Supply						
BUS FREQUENCY	Number of buses serving a bus stop between 6am and 10am on Tuesday within 400 meters of a TZ centroid	Thousands	0.19	0.15	0.02	0.77
Land use density						
POPULATION DENSITY	Population within 800 meters of a TZ centroid	Thousands	22.08	5.59	11.45	33.15
Land use diversity						
LANDMIX	Entropy of land use mix	n/a	0.13	0.01	0.09	0.17
Land use design						
PSEUDO NODES[1]	Number of pseudo nodes within 800 meters of a travel zone centroid	Thousands	1.36	0.62	0.76	4.14
Accessibility						
DISTANCE TO PT STOP	Distance between households and the nearest train station or bus stop	meter	0.24	0.08	0.12	0.59
PT STOPS	Number of train stations and bus stops within 800 meters of a household	n/a	41.45	7.58	25.60	60.77

[1]See Appendix for further discussion on the pseudo nodes

4.2 Model estimation

Table 5 summarises the estimation results of the dynamic PAM using the Ordinary Least Squares (OLS) estimator using a double log functional form. Based on the adjusted R-squared and the Ramsey's RESET test, the double-log model outperforms other functional forms which were investigated and shows no significant omitted variable in contrast to all other models where the omitted variable bias was identified as significant. The double-log model has no heteroscedasticity or autocorrelation present, evident in both the linear model and the linear-log models which were also estimated. The double-log model also demonstrates better explanatory power measured by the significance of the explanatory variables.

All parameters that are significant show the expected signs. The parameter of the lagged dependent variable at 0.245 suggests that if public transport demand in the previous period was to increase by ten percent, then current public transport demand would increase by 2.45 percent change in the current period. Higher prices, higher income, greater age, and more pseudo nodes have negative impacts on public transport demand, whereas greater bus frequency and higher population density have a positive impact. Land use mix and accessibility measures, in terms of distance and the number of public transport stops are insignificant, possibly as a result of the cohort level aggregation which reduces the variation of these measures, and possibility because of the high correlation between bus frequency and accessibility measures.

Table 5. Estimation Results of the Double-log Partial Adjustment Model

Parameter	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
LAG1	0.245	0.067	3.65	0.000	0.113	0.378
PRICE	-0.219	0.076	-2.89	0.004	-0.368	-0.070
INCOME	-0.160	0.062	-2.57	0.011	-0.282	-0.037
AGE	-0.573	0.086	-6.63	0.000	-0.743	-0.403
BUS FREQUENCY	0.148	0.051	2.90	0.004	0.048	0.249
POPULATION DENSITY	0.596	0.152	3.93	0.000	0.297	0.895
LAND MIX	-0.028	0.121	-0.24	0.814	-0.267	0.210
PSEUDO NODES	-0.458	0.109	-4.22	0.000	-0.673	-0.244
DISTANCE TO PT STOP	0.068	0.066	1.03	0.303	-0.062	0.198
PT STOPS	-0.174	0.151	-1.16	0.249	-0.470	0.123
CONSTANT	-0.164	1.645	-0.10	0.921	-3.406	3.078
Observations	236					
R-squared	0.877					
Adjusted R-squared	0.872					
Ramsey RESET Test						
Prob > F	0.072					
Breusch-Pagan Test						
Prob >Chi^2	0.074					
Wooldridge test						
Prob > F	0.423					

The short-run ($\bar{e}_k^{SR}$) and long-run ($\bar{e}_k^{LR}$) demand elasticities with respect to each of the significant variables are estimated based on the parameters using equation (7), with results presented in Table 6.

$$e_k^{LR} = \beta_k / 1 - \lambda \tag{7}$$

Where
β_k: parameter of variable k
λ: parameter of the lagged dependent variable

The short-run and long-run price elasticities estimated from the dynamic model are -0.22 and -0.29 respectively, suggesting that a ten percent increase in price is expected to reduce public transport demand by 2.2 percent in the short run, but it will reduce public transport demand by 2.9 percent in the long run. The age elasticity is -0.57 in the short run and -0.76 in the long run. The age elasticities appear to be high since a one-hundred percent increase in age gives a significant change over a life cycle. For example, students aged around 20 years old with high public transport demand will become middle-age people in the workforce after a one-hundred percent increase in age, who are expected to have a lower usage of public transport in the context of Sydney.

The two land use variables, population density and pseudo nodes, also have moderately high elasticities because a one-hundred percent change in population density and pseudo nodes indicates a dramatic changes in land use, so population density and number of pseudo nodes have strong impacts on public transport demand in terms of their percentage change, and the magnitude of impacts are greater than price, income, and bus frequency.

Table 6. Short-run and Long-run Demand Elasticities

	Dynamic Model	
	Short-Run	Long-Run
PRICE	-0.22	-0.29
INCOME	-0.16	-0.21
AGE	-0.57	-0.76
BUS FREQUENCY	0.15	0.20
POPULATION DENSITY	0.60	0.79
PSEUDO NODES	-0.46	-0.61

4.3 Demand forecasting

Using a multivariate direct demand model with explanatory variables means that the first step of demand forecasting is projecting the explanatory variables as predictors into the future. As the PAM is estimated using data from 1997 to 2009, 2009 is selected as the base year for forecasting future demand. The predictors are projected for 2010 and 2011 and then to 2026 in five year intervals using various data sources as summarized in Table 7. Public transport demand is forecast for 2011 to 2026 in order to be compared with population forecast years and Australian Census years with the same timeframe. Public transport demand in 2010 is also forecast using the demand model given that the number of public transport trips has been observed and published, so a comparison between the observed demand and forecast demand in 2009 and 2010 can be used as another approach to assess the accuracy of the forecast model.

The projection of public transport price uses the Urban Transport Fare Index of New South Wales, published by Australian Bureau of Statistics (ABS) (2012b). This index is a subgroup of the Consumer Price Index (CPI) for which historical data are also available. As there is no specific methodology for predicting future public transport price, the historical average percentage increase in Urban Transport Fare Index from 1997 to 2009 is used as the average annual price change (1.03 percent per year) for the forecast years, with all indices being adjusted to real terms based on 1997 CPI.

Table 7. Projections of Predictors for Demand Forecasting

Variable	Annual % change	2009-2026 total change	Data Source
PRICE	1.03%	19%	ABS (2012b)
INCOME	1.90%	38%	ABS (2012a)
AGE	0.50%	9%	ABS (2008)
BUS FREQUENCY	0.90%	16%	BTS (2012c)
POPULATION DENSITY	1.40%	24%	BTS (2012d)
PSEUDO NODES	0%	0%	Assumed to be time-invariant
LAND MIX	0%	0%	Assumed to be time-invariant
DISTANCE TO PT STOP	0%	0%	Assumed to be time-invariant
PT STOPS	0%	0%	Assumed to be time-invariant

Annual person income is projected forward using the historical weekly income released by ABS (2012a). This weekly income is equally weighted to the annual incomes for each year between 1997 and 2008. The historical trend shows that, on average, annual income has increased in NSW by 3.86 percent in money terms, which is slightly higher than the average increase of the Australian CPI at 2.64 percent. This is converted to an average of 1.9 percent in real terms.

The growth in the age variable is not as great as that of the income and price variables. According to the Australian Historical Population Statistics published by ABS (2008), the median age in NSW has been increasing by around 0.5 percent per year since 1998, from 35.2 years in 1998 to 36.9 in 2007. In this study this is used as the future age increase for demand forecasting.

The three variables discussed above are projected forward on the basis of historical statistics collected by ABS. These ABS statistics are based on the geography of NSW state. As no further level of aggregation is publicly available, the projections for the SGMA are assumed to be the same as for NSW.

On the other hand, bus frequency and population density are projected forward using the SGMA data forecast by the Bureau of Transport Statistics (BTS). BTS forecasts travel demand by trip mode for the SGMA using the Strategic Travel Model. Bus frequency and population density are two inputs for this Strategic Travel Model and are used in this study. The projection of bus frequency can be retrieved from Transport Supply and Demand Forecasts for the Greater Metropolitan Area published by BTS (2012c), in which the bus frequency is assumed to increase by around 0.9 percent per year between 2006 and 2036. The increase of population density is assumed to be proportional to total population growth in the SGMA, since land area is fixed over time. The population growth in the SGMA is forecast by BTS (2012d) based on 2006 Census data for each five-year interval between 2006 and 2036. This population forecast is non-linear and estimated by taking account of various factors such as the supply of dwellings, birth and deaths rates, and migration flows. Populations between each two forecast years are linearly weighted. For this study, the average population growth is estimated at 1.4 percent per year between 2009 and 2026.

Other variables including pseudo nodes, land use mix, walk distance to the nearest public transport stop, and number of bus stops are assumed to be time-invariant in the forecast model. This is because they are time-invariant variables in the pseudo panel dataset and there is no historical data or forecast data available to estimate a reasonable increase rate. However, these are subject to sensitivity tests in the next section to investigate how public transport demand might change in response to the changes in these time-invariant variables as this relevant for policy analysis.

Based on the projected variables introduced above, public transport demand in the SGMA for future years is forecast using the double-log dynamic PAM. Public transport demand is first predicted for 2009 as the base year demand, followed by 2010, 2011, and then every five years until 2026.

As the dynamic demand model defines the dependent variable as the number of public transport trips per person per day, this is aggregated to total public transport demand on an average day in the SGMA in 2009 by multiplying the predicted number of trips per person (0.336 trips in 2009) by the total population of the SGMA (5,317,330 persons in 2009). The daily public transport demand is then multiplied by 365 days to estimate the annual public transport trips in the SGMA in order to compare the annual number of public transport trips published in the Household Travel Survey (HTS) report (BTS, 2012e). Public transport demand for future years is forecast using the dynamic models based on the projected data in future years.

Table 8 summarises these forecast results and compares these to the reported demand published by BTS (2012e) and the forecast results by the ARIMA model. The reported demand published by BTS is estimated from the SHTS data by expanding the sample observations to the total population of the SGMA through a weighting scheme (2011a). This statistic is published annually and the most recent data available is in 2010.

Table 8. A Comparison of Demand Forecasting Results

Year	Reported Demand (HTS report)		Forecast Demand by PAM			Forecast Demand by ARIMA model	
	PT Trips (million)	Growth[1]	PT Trips (million)	Growth[1]	Difference[2]	PT Trips (million)	Growth[1]
2009	606.7	N/A	651.5	N/A	7.32%	N/A	N/A
2010	624.4	2.90%	659.6	1.29%	5.64%	N/A	N/A
2011	N/A	N/A	675.8	2.47%	N/A	512.30	N/A
2016	N/A	N/A	714.3	5.69%	N/A	580.40	13.29%
2021	N/A	N/A	756.1	5.85%	N/A	651.61	12.27%
2026	N/A	N/A	797.6	5.49%	N/A	722.86	10.94%

[1]As compared to the demand forecast for the previous time period (2009-2010-2011-2016-2021-2026)
[2]As compared to the reported demand based on the SHTS report published by BTS (2012e).

Comparing the reported demand and forecast demand in the base year 2009, the forecast demand is higher than the reported demand by 7.32 percent in 2009 and 5.64 percent in 2010. This difference could partly result from the weighting scheme used by BTS which is different to the aggregation process of this analysis. The forecast demand by the ARIMA model appears to be substantially lower than the reported demand in the HTS report, and this is because the HTS report includes free school bus trips which are not counted in the monthly bus boarding data estimated for the ARIMA model. The forecast demand should not be compared to the HTS report given their difference basis.

The validity of the demand forecast can also be evaluated by comparing the demand changes over time between the reported demand and forecast demand shown in Table 8. The growth rate of the reported demand between 2009 and 2010 is 2.90 percent, which is close to the growth rate of the forecast demand using the PAM of 1.29 percent. The growth rate of the forecast demand between 2011 and 2026 is around five percent to six percent for each five-year basis, whereas it is predicted to be around 10 percent to 13.29 percent growth for each five years with the ARIMA model. The five year growth rate of the ARIMA model is approximately equivalent to two percent per year, which appears to be closer to the annual growth rate of the reported demand by HTS report, suggesting a potentially stronger forecast accuracy.

4.4 Sensitivity analysis

The future is always uncertain and this applies to future changes in explanatory variables. Sensitivity analysis is therefore conducted using the PAM to forecast public transport demand based on various scenarios. These scenarios are designed by adjusting the projections of explanatory variables based on the base scenario as presented in Table 7. The results of the sensitivity analysis are summarised in Table 9. The base scenario, which assumes that all explanatory variables will change as projected in Table 7 in the future, shows that public transport demand is forecast to increase by around 26.53 percent between 2009 and 2026. This growth rate is used as a baseline to compare the following scenarios.

Table 9. Sensitivity Analysis of Public Transport Demand Forecasting

(Unit: million trips)

Year	Base Scenario	Price (+0%) +1.03%[1]	Bus Frequency (+1.5%) +0.9%[1]	Density (+2%) +1.4%[1]	Pseudo Nodes (-0.5%) +0%[1]	Combined Effect[2]
2009	672.0	672.0	672.0	672.0	672.0	672.0
2010	684.0	685.5	684.6	685.4	685.6	689.1
2011	704.3	707.8	705.7	700.6	707.9	709.2
2016	751.9	766.9	757.7	773.8	767.3	811.8
2021	801.0	829.3	812.0	850.2	830.0	924.5
2026	850.3	893.4	867.0	931.8	894.5	1,050.3
2009-2026 Total Growth	26.53%	32.96%	29.02%	38.67%	33.12%	56.30%
Increased Demand[3]	-	6.42%	2.48%	12.14%	6.59%	29.77%

[1]Percent change in the Base Scenario
[2]The scenario that combines all the scenarios on the left
[3]The total increased public transport demand as compared to the base scenario in 2026

The first sensitivity analysis assumes a constant public transport price over time, in comparison to the assumption of a 1.03 percent annual increase in the base scenario, whilst keeping the changes of other variables the same as the base scenario. This scenario is tested to examine how public transport demand changes in response to adjustments of public transport price. The results shows that public transport demand is expected to increase by 32.96 percent from 2009 to 2026 which is higher than the demand growth of the base scenario by 6.43 percent, indicating that the public transport demand could be increased by 6.43 percent if the price of public transport was to remain constant in the future as compared to the current levels of average price increase.

The next scenario investigates the extent public transport demand can be increased by providing more frequent bus services. This scenario assumes a higher increase in bus frequency at 1.5 percent per year as opposed to the 0.9 percent of the base scenario. The forecast result shows that public transport demand will increase by 29.02 percent which is slightly higher than the base scenario as expected because the long run elasticity with respect to bus frequency is positive at 0.20, and the demand increase between 2009 and 2026 is slightly lower than the first scenario as a result of the smaller elasticity as compared to the price elasticity at -0.29 in absolute terms.

The third scenario assumes an increase in population density of two percent per annum which is higher than the base scenario of 1.4 percent. This assumption is to examine the impact of increasing population density on public transport demand. A two percent annual increase in population density means that there will be a 34 percent increase from 2009 to 2026. As shown in Table 9, when population density increases at two percent annually, public transport demand is forecast to increase by 38.67 percent from 2009 to 2006 as compared to 26.53 percent increase in the base scenario. This result shows that public transport demand is more elastic to population density than with regard to public transport price or bus frequency.
Similar results can be found in the next scenario which assumes a 0.5 percent reduction in pseudo nodes. Although the number of pseudo nodes is assumed to be time-invariant in the base scenario, in the long-term it is possible to slightly change the road network by reducing cul-de-sacs and by designing grid networks in new communities to improve the walking environment and accessibility to local public transport stops. Around 8.5 percent of pseudo nodes would be reduced by 2026 as a result of an annual reduction rate of 0.5 percent. This scenario results in public transport demand increasing by 33.12 percent between 2009 and 2026 which is around 6.59 percent higher than the base scenario. This suggests that public transport demand could be effectively increased by only a slight improvement in the walking environment of the built environment and without a dramatic reform of the road network.

The last scenario, which combines all the scenarios introduced in Table 9, shows that the total public transport demand could be increased by 56.30 percent between 2009 and 2026 which is 29.77 percent higher than the base scenario. This sensitivity analysis demonstrates the way in which public transport demand can be forecast, based on various policy scenarios. The findings suggest that land use changes in terms of population density and number of pseudo nodes are expected to have a greater impact on public transport demand than changes in price or bus frequency, and the public transport demand could be increased by a total number of 56.30 percent if all the four policy scenarios in the sensitivity analysis are achieved.

4. Conclusion

This paper applies a univariate ARIMA model and a multivariate PAM to forecast public transport demand in the SGMA. The research findings contribute to the literature of conventional public transport demand forecasting studies by adding the elements of temporal effect and a comprehensive set of land use variables in the direct demand modelling studies, as well as a comparison to the time-series modelling approach.

The ARIMA model shows strong forecasting power given that the overall difference between the actual demand and forecast demand in 2011 is only around 0.67 percent. The ARIMA model is also able to control for the seasonal effects, and as a result the prediction error in the monthly forecast public transport demand ranges between -2.48 percent and 1.79 percent.

The PAM model constructs a causal statistical relationship between public transport demand and its explanatory variables, with short-run and long-run demand elasticities identified. The forecasted demand using the PAM model appears to over-estimate the public transport demand by around five to seven percent in a single year as compared to the observed demand in the HTS report. However, this difference is subject to the way in which the public transport demand is weighted to the total population of the SGMA in the HTS report and a clear measure of the actual forecasting power of the PAM is unavailable since there is no actual demand for comparison.

In terms of a comparison of the two methods, it appears that the ARIMA model demonstrates better prediction accuracy both in terms of the small differences with actual public transport boarding data and the growth rate of demand. This finding is in line with previous studies comparing the two methods for demand forecasting (du Preez and Witt, 2003; Li et al., 2010). However, the ARIMA model is not able to investigate the causal relationship and identify those exogenous factors which influence public transport demand, and hence it may not be suitable for long-term demand forecasting where these factors may change. Instead, the PAM, which captures the temporal effect of travel demand and the influence of explanatory variables on public transport demand, provides a framework in which a sensitivity analysis can examine likely future forecasts compatible with various policy scenarios. Thus, it would appear the choice of the forecasting method must depend on the primary objective of the demand forecasting. The ARIMA model should be preferred if the short-term public transport demand in the following years is of interest. In contrast, the PAM should be favoured if the objective is to understand the demand changes in response to various policy scenarios in the long term.

Appendix

This analysis uses the number of pseudo nodes within 800 meters of a Travel Zone centroid to measure land use design. Pseudo nodes are retrieved from the road network by measuring the curvature and the number of dead ends in the built environment. Figure A1 and Figure A2 illustrate pseudo nodes in two contrasting walking environments. In general, a built environment with more curvy roads and more cul-de-sacs has more pseudo nodes than an area with a grid network. The hypothesis of the relationship between public transport demand and pseudo nodes is negative, that is, public transport demand is expected to be higher in areas with fewer pseudo nodes as a result of the better walking environment. This hypothesis is based on previous studies which found that people tend to drive less and walk or use public transport more in areas with fewer cul-de-sacs since this provides better walkability and connectivity in the built environment (Cervero and Kockelman, 1997; Rajamani et al., 2003).

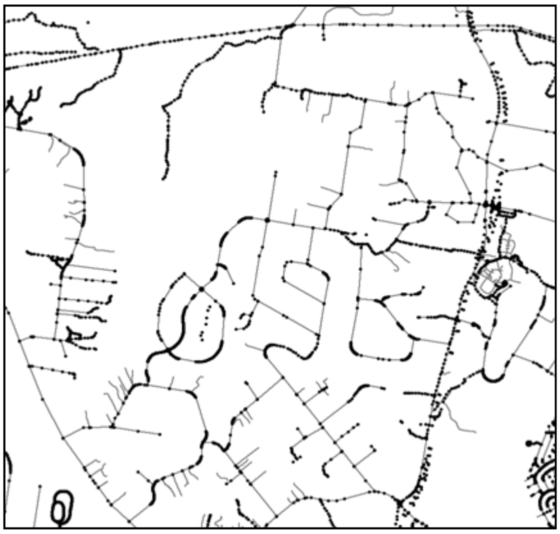


Figure A1. Pseudo nodes in a cul-de-sac built environment

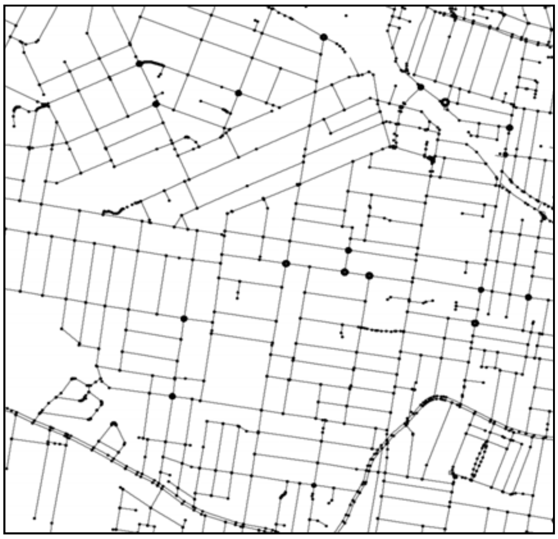


Figure A2. Pseudo nodes in a built environment with a grid road network

References

Australian Buerau of Statistics, (2008). *Median age by sex, states and territories*. Australia.

Australian Buerau of Statistics, (2012a). *Average Weekly Earnings, New South Wales (Dollars) - Original*. Australia.

Australian Buerau of Statistics, (2012b). *CPI: Group, Sub-group and Expenditure Class, Index Numbers by Capital City*. Australia.

Department for Transport, (2012). *Rail Passenger Demand Forecasting Methodology*. the United Kingdom, Available online at: http://www.dft.gov.uk/webtag/documents/expert/unit3.15.14.php.

Babcock, M. W., Lu, X., and Norton, J., (1999). Time series forecasting of quarterly railroad grain carloadings. *Transportation Research Part E*, 35, 43-57.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., and White, P., (2004). T*he demand for public transport: a practical guide*. TRL Limited: Transport Research Laboratory, Wokingham, the United Kingdom.

Blainey, S. P., and Preston, J. M., (2010). Modelling local rail demand in South Wales. *Transportation Planning and Technology*, 33, 55-73.

Box, G. E. P., and Jenkins, G. M., (1970). *Time series analysis: Forecasting and control*. Holden Day, San Francisco, the United States.

Buerau of Transport Statisitcs, (2011a). *2009/10 Household Travel Survey Summary Data*. Sydney, Australia.

Buerau of Transport Statisitcs, (2011b). *Overview of the Sydney Strategic Travel Model*, Sydney, Australia.

Buerau of Transport Statisitcs, (2012c). *Statistics in TransFigures: Travel Forecasts 2006-2036*. Sydney, Australia.

Buerau of Transport Statisitcs, (2012d). *Summary Population Forecasts 2006-2046*. Sydney, Australia.

Buerau of Transport Statisitcs, (2012e). *Summary Transport Statistics by Region*. Sydney, Australia.
Burger, C. J. S. C., Dohnal, M., Kathrada, M., and Law, R., (2001). A practitioners guide to time-series methods for tourism demand forecasting — a case study of Durban, South Africa. *Tourism Management*, 22, 403-409.

Cervero, R., (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D*, 7, 265-284.

Cervero, R., (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. *Journal of the American Planning Association*, 72, 285-295.

Cervero, R., and Kockelman, K., (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D*, 2, 199-219.

Chatfield, C., (1989). *The Analysis of Time Series: An Introduction*. Fourth Edition, Chapman and Hall, London, the United Kingdom.

Dargay, J., (2007). The effect of prices and income on car travel in the UK. *Transportation Research Part A*, 41, 949-960.

Dargay, J., Clark, S., Johnson, D., Toner, J., and Wardman, M., (2010). A forecasting model for long distance travel in Great Britain. *12th World Conference on Transport Research*, Lisbon, Portugal.

Dargay, J. M., (2001). The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A*, 35, 807-821.

Dargay, J. M., and Hanly, M., (2002). The demand for local bus services in England. *Journal of Transport Economics and Policy*, 36, 73-91.

Dargay, J. M., and Vythoulkas, P. C., (1999). Estimation of dynamic car ownership model: a pseudo-panel approach. *Journal of Transport Economics and Policy*, 33, 287-302.

Deaton, A., (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30, 109-126.

du Preez, J., and Witt, S. F., (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19, 435-451.

Ediger, V. Ş., and Akar, S., (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 35, 1701-1708.

Graham, D. J., Crotte, A., and Anderson, R. J., (2009). A dynamic panel analysis of urban metro demand. *Transportation Research Part E*, 45, 787-794.

Hunta, Ü., (2003). Forecasting of railway freight volume: Approach of Estonian railway to arise efficiency. *Transport*, 18, 255-258.

Kennedy, D., (2013). *Econometric models for public transport forecasting*. NZ Transport Agency, Wellington, New Zealand.

Kitamura, R., (1990). Panel analysis in transportation planning: an overview. *Transportation Research Part A*, 24A, 401-415.

Li, Z., Rose, J. M., and Hesher, D. A., (2010). Forecasting automobile petrol demand in Australia: An evaluation of empirical models. *Transportation Research Part A*, 44, 16-38.

Lim, C., and McAleer, M., (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, 23, 389-396.

Ljung, G. M., and Box, G. E. P., (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.

Makridakis, S., and Wheelwright, S. C., (1989). *Forecasting Methods for Management*. Fifth Edition, John Wiley and Sons, Canada.

Owen, A. D., and Philips, G. D. A., (1987). An econometric investigation into the characteristics of railway passenger demand. *Journal of Transport Economics and Policy,* 21, 231-253.

Preston, J., (1991). Demand forecasting for new local rail stations and services. *Journal of Transport Economics and Policy*, 25, 183-202.
Rajamani, J., Bhat, C. R., Handy, S., Knaap, G., and Song, Y., (2003). Assessing the impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. *Transportation Research Record*, 1831, 158 - 165.

Rodriguez, D. A., and Joo, J., (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D*, 9, 151-173.

Tsai, C. H., and Mulley, C., (2013 Forthcoming). Identifying short-run and long-run public transport demand elasticities in Sydney: a pseudo panel approach *Journal of Transport Economics and Policy*, Available online at http://www.ingentaconnect.com/content/lse/jtep/pre-rints/tsaiandmulley.

Voith, R., (1991). The long-run elasticity of demand for commuter rail transportation. *Journal of Urban Economics*, 30, 360-372.

Wardman, M., (2006). Demand for rail travel and the effects of external factors. *Transportation Research Part E: Logistics and Transportation Review*, 42, 129-148.

Wardman, M., and Tyler, J., (2000). Rail network accessibility and the demand for inter-urban rail travel. *Transport Reviews*, 20, 3-24.

Weis, C., and Axhausen, K. W., (2009). Induced travel demand: evidence from a pseudo panel data
based structural equations model. *Research in Transportation Economics*, 25, 8-18.