Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Using census data to predict commuter bus usage in Adelaide's bus catchment areas

Sekhar Somenahalli

STEM-UniSA, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA 5095

Email for correspondence: sekhar.somenahalli@unisa.edu.au

1. Introduction

Accessibility, quality of service, and congestion conditions were the main factors influencing public transport (PT) patronage in Australia, although fares also played a significant role in some cities (Biermann et al., 2020; Pund 2002). However, due to the flat fare structures in Adelaide, PT patronage is less likely to be affected by fares. In Australian cities, commuter travel is not the only form of travel, but it is the most significant single contributor to traffic volume, especially during peak hours. It is, therefore, essential to study commuter travel patterns.

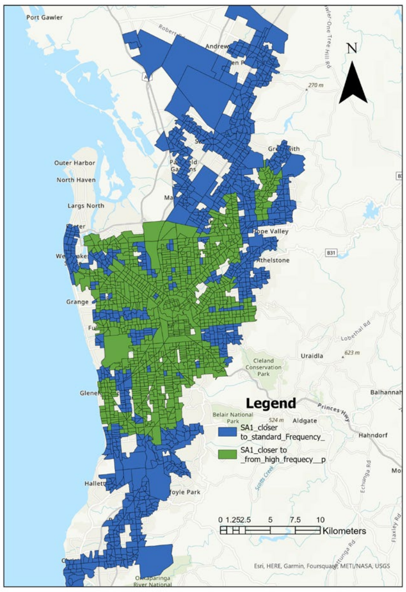
The public transport system in medium-sized cities such as Adelaide, South Australia, plays two major roles. The first is to accommodate choice riders or those who choose to use public transport for their trip-making despite having access to a motor vehicle. These commuters choose public transportation over other modes due to traffic congestion or high parking costs at work, located in most small cities within Central Business Districts (CBDs). In peak work travel periods, choice riders heavily use public transportation (Clifton and Mulley, 2016; Hensher et al., 2020). Other major roles of public transport include providing basic mobility for those segments of the population who are too young, old, or otherwise unable to drive due to physical, mental, or financial difficulties. The study will focus on those choice riders who use bus transportation for work travel. Public transport activity in Australia is increasing slightly, but it is not keeping pace with the growing demand for car travel (Hensher,1998). As part of transport planning, commuter data is extremely helpful since it provides information about utilising public transport options. According to Newman and Kenworthy (1996), residential dispersion in Australia has led to complexity in commuting patterns underscored by increasing car dependency. During this study, the overarching objective is to determine whether bus commuter trips are influenced by socioeconomic and other factors, such as bus service level of accessibility and distance to the city centre, to determine whether these relationships can be used to predict their trips. This study aims to predict bus usage in Adelaide's bus catchment areas using 2021 census data with the help of multiple regression and geographically weighted regression models.

2. Study area and data for this research

Adelaide is the capital and most populous city of South Australia and the fifth-largest city in the country. This research focuses on greater Adelaide, excluding Gawler and the hills. This analysis uses Statistical Areas Level 1 as the small spatial unit. An analysis of 2021 census data forms the basis of this study. Despite its potential, census data on work journeys has been underutilised. For each statistical area level 1, bus usage data is extracted from these journey-to-work data tables, and relationships are established with various explanatory variables that represent socioeconomic data and other variables such as distance to the CBD, accessibility and bus service.

These statistical areas are grouped into two main categories based on accessibility to bus transport and level of service. Transport networks are categorised according to their functional hierarchy. The functional hierarchy of the Land Transport Network in South Australia identifies corridors that are important for different modes of transportation. The first step is to separate high-frequency bus corridors from standard-frequency bus corridors, identify all bus stops within those corridors, and create 400-meter buffers to account for accessible bus stops. Then, all the SA1 that intersect buffers are shortlisted for further analysis. For the development of models, we use two sets of statistical areas, one located 400 meters from high frequency and one located 400 meters from standard frequency.

Figure 1: High-frequency and standard-frequency bus corridor catchment SA1s



Earlier studies have shown that the distance of the road network from the CBD (Central Business District) is directly related to public transport patronage. The network analysis in ArcGIS Pro was used to extract this information for each SA1. Each SA1 centroid was derived by converting the 'feature to point' tool, then a road network data set was created for the greater Adelaide region, and OD (Origin and Destination Distance Matrix) was generated from the CBD on all SA1 centroids.

3. Methods

Based on the Journey to Work census data for 2021, both nonspatial and spatial regression techniques were used at the statistical area level 1 (SA1) to understand and estimate bus transport demand. In the first step, the explanatory variables were sorted using exploratory regression techniques to conduct regression analysis using Ordinary Least Squares (OLS) regression. A nonspatial regression model, OLS, assumes that the relationship between the dependent and explanatory variables is constant across space. GWR analyses subsequently used the same explanatory variables (from the OLS model). For each feature in the dataset, GWR constructs a local regression equation. Local multicollinearity occurs when residual values for an explanatory variable cluster spatially. OLS and GWR models were developed for two groups of SA1s (the SA1s on high-frequency bus corridor catchments and the SA1s on standard-frequency bus corridor catchments).

4. Data analysis & results

4.1. Selection of variables for the Exploratory regression model

Before OLS regression analysis, all the logical and easily available explanatory variables were shortlisted, and an exploratory regression analysis was carried out to select the best-performing variables.

Table 1 below shows the explanatory variables that were shortlisted for the spatial regression models. A dependent variable was the number of people (by gender) who used buses (including one method, two methods, or more than three methods) to commute to work.

Table 1: List of Independent variables for regression models

S.No.	Explanatory variables
1	Number of people (by gender)
2	Number of people born outside Australia (by gender)
3	Number of people attending education institutions between the ages of 15 and 24 (by gender)
4	Median Personal income
5	Median family income
6	Median household income
7	Median age
8	Number of dwellings without motor vehicles
9	Number of dwellings with two or more motor vehicles
10	Total dwellings
11	Number of persons above 15 years of age (by gender specific)
12	Number of full-time employees (by gender)
13	Number of part-time employees (by gender)
14	The number of unemployed (by gender)
15	Number of people who went to work (by gender)
16	Distance to CBD

4.2. Models for high-frequency bus service corridor SA1s

4.2.1. Ordinary Least Squared (OLS) Model for bus commuter usage

Initially, the exploratory regression tool was run with bus users at the statistical local area level1 (SA1) within the High-Frequency Bus corridor as the dependent variable and all the shortlisted variables shown in Table 1 as potential exploratory variables. We were able to determine whether the candidate explanatory variables yielded any properly specified OLS models using this tool.

Three variables were shortlisted after carefully examining parameters such as minimum acceptable adj R Squared, maximum coefficient p-value cutoff, maximum VIF value cutoff, minimum acceptable Jarque Bera p-value, and minimum acceptable spatial autocorrelation p-value. They are 1) the number of overseas-born persons, 2) the number of dwellings with two or more motor vehicles 3) the number of part-time workers.

Table 2 shows the statistically significant variables (p-value less than 0.05). A significant coefficient is indicated by an asterisk next to the probability. Small probabilities are better (more significant) than those that are large. Explanatory variables that were redundant were avoided. A model with redundant variables has an overcount of variables, indicating a bias; multicollinearity is the term used to describe this. The Variance Inflation Factor test or VIF measures multicollinearity. VIF values should be less than 7.5, but the smaller, the better. In Table 2, all variables have VIF less than 7.5, indicating no redundancy.

Table 2: Summary of OLS results -Bus commuter users in high-frequency bus service corridor SA1

Variable	**Co-efficient**	**Std Error**	**Robus_t-statistic**	**Robust probability**	**VIF**
Intercept	-0.270945	0.641181	-0.194814	0.845573	-------
Overseas born persons	0.0565575	0.04429	9.146153	0.000000*	2.366633
number of dwellings >= 2 vehicles	-098944	0.010845	-5.283223	0.000000*	3.016867
number of part-time workers	0.168734	0.05137	8.953211	0.000000*	5.011320

Table 3: OLS Diagnostics

Number of observations = 1003		
Akaike's Information Criterion (AICc)	6945.703	
Multiple R-Squared value	0.510834	
Adjusted R-Squared value	**0.509265**	Probability
Joint F-Statistic	347.7508	0.000000*
Joint Wald Statistic	363.5416	0.000000*
Koenker (BP) Statistic	206.5897	0.000000*
Jarque-Bera Statistic	383.1895	0.000000

Table 3 above shows that the Koeker test is statistically significant. This indicates that the relationship between some or all of the explanatory variables and the dependent variable is non-stationary. The variables might be strong predictors of bus usage in some SA1s but weak in others. By using Geographically Weighted Regression modelling methods, it is likely that this model results can be improved. A spatial autocorrelation tool must be used before this step to verify that residuals from the OLS model are clustered.

4.2.2. Autocorrelation- Moran's I Index on residual of OLS model

In geographical space, spatial autocorrelation statistics measure and analyse dependency among observations. Using ArcGIS' Spatial Autocorrelation (Global Moran's I) tool, we measured spatial autocorrelation based on feature locations and values. Using GWR modelling, AIC values can be reduced, model strength can be increased, and local variables can be predicted. A Moran's I test showed that the residuals are clustered given that the z-score is 35.88; there is less than a 1% probability this pattern could be random.

4.2.3 Geographically Weighted Regression Mode (GWR) model

A geographically weighted regression (GWR) model was then run using the same exploratory variables as the OLS model. Table 4 shows adjusted R2 and AIC values for the GWR model. By allowing the explanatory variable relationships to vary across the study area, such as the GWR model, the adjusted R squared increased to nearly 73% in OLS. The AIC value has been reduced from 6945 (OLS model) to 6423 (GWR model), which is a good sign. GWR also produces a residual map (Figure 4), which displays over- and under-predictions. For a well-specified regression model, over- and under-predictions will be randomly distributed.

Table 4: GWR model results

Akaike's Information Criterion (AICc)	6423.845
R squared	0.7529
Adjusted R-Squared	**0.7245**

4.3. Models for standard frequency bus service corridor SA1s

After applying the same criteria as in section 4.2.1, three variables were shortlisted. They are 1) the number of overseas-born people, 2) the distance to the CBD and 3) the number of people who went to work. In the process of developing the OLS model, a few interesting points were observed. The adjusted R squared for this model was less than 0.5 (0.43), and the exploratory variables that showed significance were different. Along with 'Overseas born persons', variables such as 'distance to CBD' and 'number of people who went to work' were significant. The Koenkar BP test showed significance, so the autocorrelation tool was run. The Morans I index results again showed that the OLS model's results were clustered (with a z value of 36.875). It was, therefore, necessary to run the GWR model. It was encouraging to see that when the GWR model was run using the same exploratory variables as in the OLS model, the R squared went up from 43% to 63%.

5. Discussion

Although it is quite logical to assume that overseas-born people (perhaps recent migrants) tend to use bus services more often, it is surprising that this variable was the most important for predicting bus commuter trips in both high-frequency and high-frequency and standard-frequency bus corridor catchments. Similar observations were also made by Allen et al.,2021. In the next few years, migrants are expected to increase in number, so it is essential to observe where they are settling, which will assist in attracting higher bus patronage.

It has been found that part-time workers have shown a positive relationship with bus commuter usage in high-frequency bus catchment areas, even though these trips may occur during off-peak times. The higher levels of service in these catchment areas, even during off-peak hours, might also have contributed to this positive relationship. As a result, this variable was not significant in the standard frequency bus corridors, perhaps due to inadequate bus service levels.

Lastly, high vehicle ownership, a proxy for income, negatively impacts bus patronage in high-frequency bus corridor catchments. In standard frequency bus corridor catchments, neither income nor vehicle ownership had a significant impact. Instead, 'distance to CBD' was negatively correlated with bus commuter usage.

6. Conclusions

The results indicate that variables derived from census data are highly significant in modelling bus commuter trips. This study shows that contrary to the view that socioeconomic factors have a limited effect in estimating mode of travel models, they have a significant effect when combined with bus service levels and urban form variables. It could be because models were built after separating statistical areas into those in catchment areas of high-frequency and standard-frequency bus service corridors. As a result, the bus level of service component has been addressed. The addition of urban form variables, such as distance to CBD, resulted in building an acceptable model (above 0.5 R squared) using socioeconomic data from the census. Finally, using geographically weighted regression models, we can significantly improve the strength of spatially varying relationships (i.e. R squared increased from 0.53 to 0.72).

Using these models, bus route planning agencies can identify new bus routes and upgrade standard-frequency bus corridors into high-frequency ones.

References

Allen, J., Farber, S., Greaves, S., Clifton, G., Wu, H., Sarkar, S. & Levinson, D. M. (2021) Immigrant settlement patterns, transit accessibility, and transit use. Journal of Transport Geography, 96, 103187.

Biermann, S., Smith, B., Olaru, D., Cardell-Oliver, R. & Reed, T. (2020) Factors Influencing Public Transport Patronage Trends Perth 2015 to 2019. Perth: Planning and Transport Research Centre, University of Western Australia.

Clifton, G. T. & Mulley, C. (2016) A historical overview of enhanced bus services in Australian cities: What has been tried, what has worked? Research in Transportation Economics, 59, 11-25.

Hensher, D. A. (1998) The imbalance between car and public transport use in urban Australia: why does it exist?1An earlier version was presented at the 1998 Annual Conference of the Australian Bus and Coach Association, Cairns, 26–29 April.1. Transport Policy, 5(4), 193-204.

Hensher, D. A., Wong, Y. Z. & Ho, L. (2020) Review of bus rapid transit and branded bus service network performance in Australia. Research in Transportation Economics, 83, 100842.
Newman, P. W. G. & Kenworthy, J. R. (1996) The land use—transport connection: An overview. Land Use Policy, 13(1), 1-22.

Pund, G. (2002) Urban form, level of service and bus patronage in Eastern Sydney, 25th Australasian Transport Research Forum. Canberra, 2 Oct 2002 Bureau of Transport and Regional Economics.

Walks, A. (2015) The Public City: Essays in honour of Paul Mees. Planning Theory & Practice, 16(4), 597-599.