Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Neighbourhood typology and bus use: A simplified approach to predict bus demand

Jeyashivraj Parthiban[1], Laura Aston[2], Graham Currie[3]

[1]Built and Natural Environment at Wellington Shire Council, 18 Desailly Street (PO BOX 506), Sale VIC 3850

[2] Arup, Sky Park, One Melbourne Quarter, 699 Collins Street, Docklands VIC 3008

[3]Public Transport Research Group, Institute of Transport Studies, Department of Civil Engineering, Monash University, Wellington Road, Clayton VIC 3800, Australia

Email for correspondence: jeyashivrajp@wellington.vic.gov.au

Abstract

To compete with a growing number of alternative mobility options, bus operators and governments must be able to accurately anticipate bus demand and plan accordingly to encourage greater use in diverse areas. There is general understanding of the network design principles and built environment attributes that are more conducive to public transport use compared to private automobile use. However, relatively little is known about the importance of the built environment for bus use. Furthermore, there are some practical barriers to developing accurate and flexible bus demand prediction tools.

This study addresses these two gaps by developing a bus demand prediction model based on neighbourhood typologies. The built environment attributes of bus stop catchments are combined, and clustered according to similar attributes. We compare the prediction performance of the model that predicts ridership using typologies, to a conventional multivariate model with individual built environment variables.

The typological model explains slightly less variance but offers simpler interpretation and is more generalisable. Prediction models for individual neighbourhood typologies suggests that the relationships between the built environment and bus demand differ in parts of a city with different built environments. When the purpose of a model is to yield the most reliable prediction, the typological approach offers a simple way to predict demand while capturing spatial variation in the built environment. However, in situations where it is of interest to identify appropriate interventions for a particular site, it may be appropriate to collect and examine data for similar locations only.

1. Introduction

In many world cities, bus ridership is undergoing long-term decline (Berrebi & Watkins 2020). In Australia, bus use is growing slower than other public transport modes, namely train and tram, while its mode-share is decreasing (Pemberton 2020). Furthermore, the ubiquitous disruption caused by the COVID-19 pandemic has exacerbated this decline by supressing daily travel and motivating a mode-shift away from shared transportation (Beck et al. 2020; de Haas et al. 2020).

Much research in the field of travel behaviour and forecasting has been directed towards understanding the determinants of transit use. The built environment is one important factor that affects public transport demand. Although some research attention has focused on this relationship for bus, there is limited uptake of prediction tools for bus that incorporate land use and urban design.

One barrier to developing effective demand prediction tools is the complexity of the built environment relationship with public transport use. Econometric models are used in research to quantify relationships. The outputs of these models can in turn be used to predict future demand and compare built environment scenarios. However, these models have become increasingly sophisticated to accommodate more variables, their interdependence; and nonlinear spatial impacts. Such advances can generate robust and precise findings for a given set of model assumptions, which may have limited application in practice.

This study addresses these gaps with its aims, to:
1. Examine the relationship between the built environment and bus use in different parts of a city.
2. Develop and evaluate a simplified approach for estimating the underlying demand for bus using built environment typologies.

In the remainder of this paper we first present recent developments and gaps in the built environment and public transport use literature. We describe the variables and built environment typologies developed for this study. We explain the analysis approach and then present the results of different models. Finally, we compare findings for bus in different parts of the city and evaluate the performance of the typological approach.

2. Literature Review

Public transport is underperforming compared to other modes of transport in many world regions (Berrebi & Watkins 2020). In addition, COVID-19 has created an imperative for public transport agencies to attract riders back following suppressed ridership due to health concern and restrictions to mobility (Tirachini & Cats 2020). Bus transport also faces the most significant threat from flexible-route shared mobility options, with which it may compete for first- /last-mile services to fixed route public transport. Bus is also more commonly found in low density areas where the potential for value capture from adjacent land use does not justify the investment in fixed guideway infrastructure. In this context, it is important to have a detailed and nuanced understanding of the determinants of bus ridership.

2.1 Factors influencing demand

A significant body of public transport research focuses on the determinants of demand. Aside from service quality, several studies suggest that public transport demand is primarily affected by sociodemographic variables (Ewing & Cervero 2010); including bus use specifically (Berrebi & Watkins 2020). Mode choice studies have also identified an important role for the importance of the built environment on public transport use (Boulange et al. 2017). The influence of stop/station catchment-area built environment on ridership is well-documented (Aston et al. 2020b; Cervero 2007; Ibraeva et al. 2020).

While the impact of service quality on public transport use is consistently positive, the influences of built environment and sociodemographic factors are less generalisable and not always linear (Holz-Rau & Scheiner 2019). Voulgaris et al. (2017) showed that neighbourhood built environment qualities synergise to produce significant impacts on mode choice in a comprehensive study spanning the USA.

2.2 Approaches to quantifying demand relationships

Public transport demand studies typically focus on measuring the impedance (cost) or benefit of travelling by different modes and adopt statistical models to estimate relationships (Maat et al. 2005). The limitations of this approach are recognised, and include unreliable estimates of demand (Flyvbjerg et al. 2005; Voulgaris 2019), a lack of nuanced understanding of user-related variables (Blitz & Lanzendorf 2020; Vecchio & Martens 2021) and findings that are not readily transferrable to policy (Dill et al. 2014). Such shortcomings are partly attributable to the simplification of the complex inter-relationships between land use and transport (Holz-Rau & Scheiner 2019). Furthermore, relevant studies suggest that relationships between the built environment and travel behaviour may vary spatially (Ding et al. 2021)

To address this, some studies have turned attention to examining overall neighbourhood character in relation to public transport use (Higgins & Kanaroglou 2016; Kamruzzaman et al. 2014; Voulgaris et al. 2017). However, these studies examine public transport use, or more specifically, rail use. Most studies examine bus ridership at the level of an administrative area (De Gruyter et al. 2020; Taylor et al. 2009; Voulgaris et al. 2017), while those that examine station-level ridership typically focus on public transport in general or rail-based transport (Higgins & Kanaroglou 2016; Rodríguez & Kang 2020). Relatively few studies have considered the built environment determinants of bus use at the stop level, preferring the route, network or larger urban areas as the unit of study (Currie & Delbosc 2011). Such units of analysis provide limited opportunity to capture the effect of neighbourhood level characteristics. Furthermore, the majority of studies include individual built environment attributes as explanatory factors, rather than overall neighbourhood character.

Of the studies that examine the impact of overall neighbourhood character, only one goes on to compare the predictors of ridership among different types of neighbourhood. Higgins and Kanaroglou (2016) found that stations with high density, higher walkability, and better mixed land use improved rates of public transport use. The analysis provided a series of transit-oriented development (TOD) planning tools tailored to station-area context. This approach responds to the growing need for context-sensitive planning tools, that accommodate differences in intra-urban context, in terms of both spatial and user variables (Taylor et al. 2009; Vecchio & Martens 2021). Stojanovski (2018) investigated built environments and bus use in Karlstad, Sweden, and identified a clustering of both types of variables into intervals which could be interpreted as neighbourhood typologies. The paper suggests that neighbourhood typology may be a simpler way of representing land use than individual variables when developing bus demand models. Further analysis would be useful to evaluate the predictions of bus use that arise using multivariate models with neighbourhood typologies instead of individual variables.

3. Method

3.1 Research setting

This study examines the relationship between bus use and the catchment built environment in Greater metropolitan Melbourne. Melbourne occupies almost 10,000km^2 and is home to over 5 million inhabitants. Land use intensity declines radially from its strong core central business district (CBD). The inner suburbs are dense both by populations and dwellings featuring multimodal transport corridors and major activity centres (Fuller & Crawford 2011). The outer suburbs that developed after the introduction of cars are less dense, primarily served by freeways with some access to public transport and extending up to 50km from the city centre (Fuller & Crawford 2011). Melbourne also has an extensive multimodal public transport network. There are 18,000 bus stops in Melbourne, compared to approximately 2,000 tram and 200 train stops. Melbourne's bus network is extensive, a large share of which operate in a 'tailor-made', meandering configuration that services the full gamut of Melbourne's diverse urban form (Pemberton 2020). Together with the diversity of the land use it services, this makes Melbourne an ideal case study for exploring the importance of neighbourhood character on bus use.

3.2 Variables

This study leverages an existing database containing built environment and sociodemographic data for Melbourne's public transport catchments (Aston et al. 2020a). This study uses four types of variables. The dependent variable is public transport ridership, measured as the average normal weekday daily boardings. Independent variables span three further themes: public transport supply, bus station-area demographics and bus station-area built environment. A complete list of variables is provided in Table 1, which also includes descriptive statistics for each cluster. A brief description of the variables is provided below, but more detailed source information and processing steps are detailed in Aston et al. 2020b.

The variables were measured according to the unit of analysis that is relevant for behaviour. Public transport supply is gathered for the services that use any of the stops in a facility, as well as those that overlap within reasonable transfer distance (120m radius of the stop). Bicycle and car parking supply was also gathered for this transfer zone. Sociodemographic data and most of the remaining built environment variables were collected for a 400 metre network catchment of the bus stop, representing the walkable access and egress catchment (Boulange et al. 2017). Sociodemographic variables captured ethnicity, education, household size and employment status. Catchment built environment variables include population, employment and commercial density, land use mix, daily living destinations reachable, housing diversity, distance to activity centres, the proportion of urban land, and the ratio of population to housing. Finally, the regional connectivity from each stop was measured in terms of the number of jobs reachable by public transport, within 30 minutes during the morning peak. The methodology and sources for these variables have been documented earlier in Aston et al. 2020b.

These variables are interrelated in ways that can affect statistical analysis. First, public transport supply may not only influence, but also be influenced by (endogenous to), demand (Taylor et al. 2009). Second, the variables tend to co-occur, or be self-reinforcing, thus introducing collinearity into statistical models. We describe how we address these issues in subsequent steps.

3.3 Bus stop neighbourhood typologies

This study uses the typologies developed in earlier published work by Aston et al. (2020b), exploring the performance of all public transport modes around stop/station catchments. The typologies were formed using cluster analysis.

An important assumption of both cluster analysis and matching is to ensure the theoretical relevance of the variables. In the context of clustering to classify public transport stops into "like" station areas, it is primarily the access and egress catchment variables that are of interest [1]. In addition, regional accessibility (distance to activity centres and access to

[1] 800 metres was used as the service area to generate clusters for a combined sample of train, tram and bus, in the original study (Aston et al. 2020b), so that the clustering was consistent across three different modes. However,

employment), though not a property of the neighbourhood catchment, are also linked to the spatial distribution of modes due to the mono-centric nature of Greater Melbourne. Finally, although demography is not a physical property of public transport facilities, the spatial distribution of population groups also varies systematically within cities (Delbosc & Currie 2011). To control for this, it is also relevant to include the station-area demographic variables in the formation of clusters.

In preparation for robust cluster analysis, factor analysis of the built environment variables was used to identify collinear variables so that this could be minimised (Hair et al. 2014). Factor analysis was carried out in RStudio (RStudio Team 2016). The ReDas package was used to confirm the appropriateness of the variables for factor analysis, using Bartlett's test to determine whether sufficient collinearity existed among variables (Maier 2019). The factanal function in the psych package was used to conduct factor analysis based on the correlation matrix of the input variables (Revelle 2018a).

Four factors were initially specified according to the groupings of density, diversity, local and regional accessibility. Several iterations of factoring were performed, with variables removed in a stepwise fashion if they did not fit the factor solution, denoted by low factor loadings. A three-factor solution, explaining 76% of variance across the catchment areas was determined to be suitable. One variable from each factor, representing employment density, accessibility; and residential density, was included in the cluster analysis. In addition to the three factors, six variables did not fit the factor solution were all included in cluster analysis. These were: proportion commercial, land use balance, land use diversity, pedestrian connectivity, bicycle connectivity and distance to nearest activity centre.

Bottom-up (non-hierarchical) clustering was used to form clusters based on the k-means clustering algorithm. This algorithm assigns observations to one of k (pre-specified) centroids and iterates the cluster centroid such that the sum of squares of observations in the cluster is minimised (Boehmke 2019). The R package cluster was used to execute the clustering algorithm in RStudio (Maechler et al. 2021; RStudio Team 2016). K-means clustering requires pre-specification of cluster centroids. Various indices exist for calculating the optimum number of centroids, based on the similarity within groups and differences between groups. When applied to group large samples of spatial units according to land use, cluster solutions were characterised by three to seven clusters (Jeffrey et al. 2019; Kamruzzaman et al. 2014; Voulgaris et al. 2017). Given the large sample size (n = 10,631 across three modes) a range of cluster solutions were compared for the best fit. A six-cluster solution was found to be most appropriate. The spatial distribution of the six resulting clusters are shown in Figure 1.

The clusters represent neighbourhood 'types' and are distinctive both in terms of their built environment characteristics, but also in terms of their relative proximity to the central business district of Melbourne. This reflects the dominant mono-centricity of Melbourne, such that development intensity decreases radially with distance from the urban downtown area. As such, the clusters were named to reflect their spatial distribution and predominant uses. Descriptive statistics of each cluster are provided in Table 1. The Urban Core cluster (neighbourhood typology 1) includes bus stops in the central business district of Melbourne. It is distinct from the other clusters given its high scores in all the built environment variables considered in the analysis. The Inner Urban cluster (2) includes bus stops beyond the urban core cluster and up

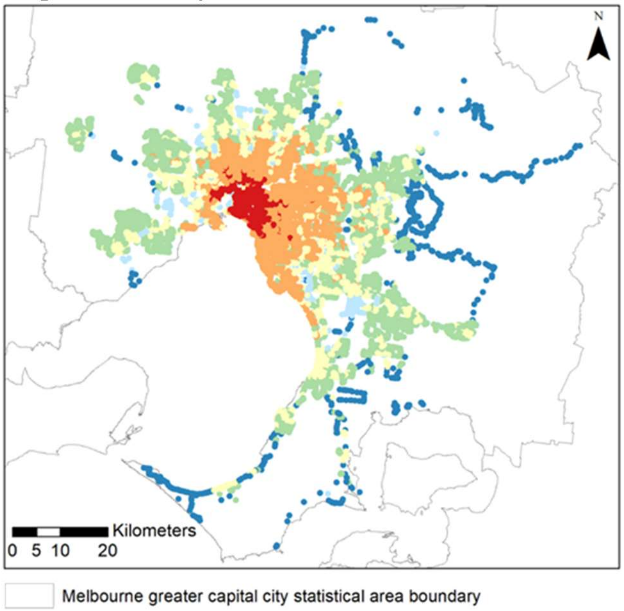
values for neighbourhood variables within the reduced bus walking catchment of 400 metres are used in subsequent econometric analysis, including this study.

to 15km from the central business district. The localities in the cluster have higher scores in all built environment characteristics than cluster 3, 4, 5 and 6. The Mixed-Use Suburban cluster (3) is relatively diverse, with mixed land uses and an abundance of multimodal transport facilities. The Residential Suburban cluster (4) is location at a similar proximity to the CBD to the Mixed-Use cluster, but scores low on most built environment attributes. The Industrial cluster (5) scores low in all built environment variables except employment access and density, signifying these are employment-oriented locations. Fringe Suburban stops (6) are located at the boundary of metropolitan, with the lowest scores for all built environment attributes.

Table 1: Averages of built environment characteristics for six neighbourhood typologies

Variable	1 Urban core	2 Inner urban	3 Mixed-use suburban	4 Residential suburban	5 Industrial	6 Fringe suburban
Cluster built environment variables						
Total patronage	101	36.6	40.5	12.9	19.8	4.48
Employment density (workers/km^2)	8,575	969	772	357	1,050	62.5
Pop. density (residents/km^2)	6,135	3,123	1,844	2,405	112	440
Commercial density	0.32	0.05	0.19	0.01	0.62	0.04
Jobs-housing balance	0.14	0.08	0.12	0.04	0.49	0.04
Land use entropy	0.48	0.31	0.47	0.23	0.30	0.37
Housing diversity	7.6	6.4	5.3	4.3	3.8	3.0
Intersection density (count per square km)	153	80.8	85.6	80.3	55.3	48.6
Bike connectivity (km)	86	318	213	142	15.6	30.6
Destination score	5.08	3.33	2.69	2.09	1.29	1.44
Bike parking cage available	0.00	0.01	0.02	0.00	0.00	0.00
Dist. To CBD (km)	3.87	13.2	23.6	28.2	22.7	44.9
Non-cluster built environment variables						
Dist. to Activity Centre (km)	0.92	1.51	1.68	2.28	3.02	9.66
% zone classified "urban" (%)	0.78	0.87	0.68	0.89	0.11	0.36
Employment access (jobs reachable in 30 minutes)	256,162	30,654	16,242	6,256	16,189	1,077
Car parking area (m^2)	48.12	102.7	222	26.7	55.8	21.1
Supply (arrivals/hour)						
Level of service (LOS)	183	116	116	76.5	76.1	44.7
LOS overlapping bus	0.32	0.32	0.34	0.27	0.23	0.21
LOS overlapping tram	0.91	0.13	0.04	0.001	0.01	0
LOS overlapping train	0.07	0.03	0.03	0.01	<0.01	0.01
LOS overlapping (total)	1.30	0.47	0.41	0.28	0.25	0.21
Socio-demographic						
Proportion full time employed	0.61	0.57	0.55	0.57	0.76	0.54
Mean household size (persons/dwelling)	2.40	2.76	2.84	5.02	5.69	5.27
Proportion of foreign-born persons	0.47	0.40	0.45	0.38	0.52	0.54
Proportion of tertiary educated persons	0.62	0.54	0.35	0.34	0.28	0.29

Figure 1: Melbourne bus stops classified by catchment built environment. Source Aston et al. (2020b)



Cluster Number	Cluster description	n (bus)
1	Urban core	319
2	Inner urban	2,844
3	Mixed use suburban	1,905
4	Residential suburban	3,548
5	Industrial	236
6	Fringe suburban	632

3.4 Regression analysis

Regression models were developed to explore relationship between different types of bus station areas, and bus use, in different parts of a city. Three models were specified for a combined sample of all bus stops. The three models comprised public transport supply variables and station-area demographics but differed in the granularity of the built-environment variables. The fourth approach segmented the sample into each of the six neighbourhood typologies, for which models were developed separately. This final approach was included to determine whether any additional useful information could be gained by examining each neighbourhood typology separately. The specification of bus stop catchment-area built environment, and sample, for each of the four approaches, is as follows:

1. Conventional (Model I): individual built variables only (all bus stops),
2. Typological (Model II): neighbourhood typology only (all bus stops),
3. Combined (Model III): individual built environment variables and neighbourhood typology (all bus stops); and
4. Segmented (Models IV – IX): individual built variables (individual models for each neighbourhood typology).

In each model, the independent variables were regressed on total patronage using ordinary least squares regression. To ensure that the values of the independent variables do not influence the results of the regression, these values were standardised before analysis. Prior to conducting each regression variables with high interdependence (collinearity), indicated by variance inflation factors (VIF) exceeding 5, were removed. A parsimonious model was formed by removing variables on at a time if they did not contribute to the explanatory power (adjusted R^2) of the model.

3.5 Addressing public transport supply endogeneity

Since public transport operators adjust supply in response to demand, it follows that in areas with higher densities, supply is likely to be higher. Efforts have been made by researchers to account for this endogeneity using instrumental variables for level of service (Diab et al. 2020; Taylor et al. 2009). We attempted to develop and instrumental model for bus use, using spatially lagged density indicators. However, these models introduced prohibitive multicollinearity into the models, so we were unable to adopt the instrumented models.

4. Results and Discussion

4.1 The impact of spatial context on bus ridership

The first aim of this study is to explore the relationship between the built environment and bus use in different parts of the city. We tested four approaches to achieve the study's second aim which was to determine the most useful and robust way of understanding bus demand in different spatial contexts.

We first ran a series of three regression models for all bus stops (9,484) in the sample. Results for Models I – III are shown in Table 2. Results suggest that the overall neighbourhood 'type' explains bus ridership almost as well as a model with individual built environment attributes. These results suggest that after accounting for individual attributes of the built environment, being in an industrial area has the largest impact on public transport use. In other words, Industrial Areas outperform other areas in terms of ridership, considering the baseline attributes of the built environment. This finding suggests that there may be latent demand for travel to industrial areas, but because their built form is not integrated with the surrounding transport and land use, usage remains low. This corroborates findings of an earlier study of Melbourne's train network, which also identified mixed use areas as those with potential for urban intensification to encourage ridership (Jeffrey et al. 2019). Conversely, Urban Core areas underperform relative to other areas, considering the built environment. Despite showing the strongest relationship with bus use in the Cluster only model, being in the Urban Core did not predict any higher ridership than being in the Fringe. This likely reflects the bias of public transport users in the Urban Core toward light rail and heavy rail, which is amply supplied there. In the combined model (Model III), Inner Urban areas were the second most strongly linked to ridership relative to the Fringe Suburban Cluster. It is followed by the residential suburban cluster, and the Mixed-Use Activity Centres.

Combining insights from Models I and III suggests that in general, higher densities and greater land use diversity, as well as active transport facilities, are linked with higher ridership. The differing relative importance of the neighbourhood typology with (III) and without (II) built environment variables in the equation suggests there is some non-linearity between the built environment and public transport use. In the next section, we unpack these relationships by typology.

Table 2: Models for bus ridership considering built environment and neighbourhood typology (Significance thresholds: * p< 0.05, ** p< 0.01 and *** p <0.001.)

		I - BE only[2]	II - Typology	III - Both
	Observations	9,483	9,484	9,484
	Variable	**Standardised (β) coefficient**		
Supply	Level of service (LOS)	0.569***	0.578***	0.564***
	LOS overlapping bus	-0.018*	-0.049***	-0.031**
	LOS overlapping train	0.081***	0.131***	0.073***
	LOS overlapping (total)		0.051***	0.015
Area-level sociodemographic variables	Proportion full time employed	0.039***	0.018*	0.028***
	Mean household size	0.029**	-0.021*	0.024**
	Proportion population foreign-born	0.143***	0.160***	0.141***
	Proportion population tertiary educated	0.080***	0.0345***	0.072***
Built environment	ln(employment density)	0.073***		0.064***
	ln(population density)	0.086***		0.105***
	Commercial density	0.056***		0.051***
	Jobs housing balance	0.024***		0.014
	Land use entropy	0.040***		0.064***
	Housing diversity	-0.048***		-0.058***
	Intersection density	0.030***		0.029***
	Destination score	0.108***		0.104***
	Bike parking cage available	0.044***		0.040***
	Dist. to CBD	0.113***		0.125***
	Dist. to Activity Centre	0.061***		0.081***
	Count of Activity Centres	0.035***		0.041***
	% zone classified "urban"	0.060***		0.039**
	Employment access	-0.019*		
	Car parking Area			0.012
Station-area typology (ref: Fringe Suburban)	Urban Core		0.280***	0.115
	Inner Urban		0.226***	0.276***
	Mixed Use Suburban		0.155***	0.136**
	Residential Suburban		0.220***	0.238***
	Industrial		0.099	0.457***
	Intercept	<0.001***	-0.193***	-0.214***
	Standard Error	1.021	0.713	0.689
	Multiple R-squared	0.525	0.493	0.527
	Adjusted R-squared	0.524	0.492	0.526
	Degrees of freedom	9,461	9,470	9,456

4.2 Predicting bus ridership by neighbourhood typology

The purpose of the analysis was to compare the variation in the effects of built environment variables on bus ridership in different clusters of Melbourne. It was hypothesised that the built environment variables affecting bus ridership and the strength of their effects would be different for each cluster. To this end, we developed ridership models for individual clusters. Results are presented in Table 3.

The second objective of this study was to identify the most important predictors of bus use for differential spatial typologies. From Table 3, it can be observed that the variables found to be significantly affecting bus ridership differ by spatial typology. Furthermore, the explanatory power of the models differed (denoted by their Adjusted R^2 values). This was not necessarily related to a cluster's size – despite being the largest cluster, the residential suburban cluster had the lowest Adjusted R^2 value. This suggests that the relative importance of all three types of variables (level of service, built environment and sociodemographics) varies with spatial context.

[2] This model includes two indicators for which new data has been used since the publication of earlier research for Melbourne. The two indicators with new data are: proportion of residents born overseas, and transit service level.

Table 3: Individual cluster models for bus ridership (Significance: * p< 0.05, ** p< 0.01, *** p <0.001.)

Regression Statistics	IV - Urban Core	V - Inner Urban	VI - Mixed Use Suburban	VII - Residential Suburban	VIII - Industrial	IX – Fringe Suburban
Observations	319	2,844	1,905	3,547	236	632
Variable	**Standardized (β) regression coefficient (significance threshold)**					
Supply						
Level of service (LOS)	0.642***	0.629***	0.597***		0.583***	
LOS overlapping bus	-0.118**	-0.055**				-0.085*
LOS overlapping train	0.072	0.136***	0.103***			
LOS overlapping (total)	0.104*	0.063**				
Area-level sociodemographic variables						
Proportion full time employed			0.026	0.052**	0.129**	-0.078
Mean household size	-0.059	0.043**	0.058**	0.027		0.065
Proportion of foreign-born persons		0.141***	0.05*	0.196***		
Proportion of tertiary educated persons		0.101***	0.054**			-0.071
Built environment						
Employment density		0.059***	0.116***	-0.093***	0.189***	-0.135*
Population density		0.063***	0.072***	0.129***		0.30***
Commercial density		0.045*	0.054**	0.03	0.123*	-0.11**
Jobs housing balance				0.06**	0.129**	0.075
Land use entropy	0.096**	0.081**	0.061***	0.113***		0.134***
Housing diversity		-0.076***	-0.02			
Intersection density			0.024	0.114***	0.211***	0.132**
bike connectivity		0.022		0.054***		0.085*
Destination score	0.137***	0.057***	0.059**	0.12***	0.065	0.253***
Bike parking cage available		0.044**	-0.066**			
Dist. to CBD	-0.201***	0.151***	0.166***			-0.24***
Dist. to Activity Centre		0.082***	0.077***	0.018		0.169**
Count of Activity Centres	0.13***	0.04**	0.076***	0.04*	0.06	
% zone classified “urban”		0.048*	0.029	0.082**		
Employment access			0.076***	0.134***		0.24***
Car parking area	0.073					
Intercept						
Multiple R-squared	0.652	0.566	0.560	0.173	0.593	0.294
Adjusted R-squared	0.641	0.561	0.557	0.170	0.578	0.277
Standard Error	1.03	1.03	0.927	1.12	0.98	1.12

Seven variables were significant in the Urban Core cluster, the most important of which was level of service, followed by proximity to the central business district. This cluster was also the strongest model, with an explanatory power of 0.641. Eighteen variables were significant for the Inner Urban cluster, and 15 were significant for the Mixed-Use suburban clusters. The Inner Urban and Mixed-Use Suburban clusters demonstrate many similar relationships. In both cases, the most important factor for ridership was level of service.

The second most important was distance from the CBD (an inverse relationship compared to that for the Urban Core). Despite being the largest cluster, the Residential Suburban cluster had the lowest explanatory power. The Fringe Suburban Cluster also had very low explanatory power. The lack of significance of level of service in these two clusters may be due to the low average level of service, with 76 and 45 departures on average between 7am and 7pm respectively, compared to almost 120 for the higher order clusters. There may be a minimum level of service threshold, below which ridership is not sensitive to frequency.

In contrast, bus stops classified as Industrial had a high explanatory power (R^2= 0.58) despite its sample size of just 236 stops. Six variables were identified to be significant, of which level of service was most important followed by intersection density. Although this cluster also had low level of service, it is distinct from Residential Suburban and Fringe Suburban in that it is predominantly employment rather than residential land. Destination score explained changes in patronage across all clusters. It can be concluded that irrespective of the part of the city in which bus is found, destination score is always an important determinant of ridership. The impact of destination score on total patronage is highest in the Fringe Suburban bus stops, and lowest in the Inner Suburban bus stops.

Employment and commercial density explain changes in total patronage across all typologies, except the Urban Core. The Urban Core had the highest employment and commercial density among all clusters. This may suggest that employment and commercial densities have reached a tipping point beyond which additional density does not attract additional riders. Conversely, employment and commercial densities had the highest impact on the Industrial typology; a logical finding considering these otherwise low-intensity neighbourhoods mostly comprise of employment opportunities.

Although segmenting demand models by neighbourhood typology improves explanatory power, there are some important limitations to this approach in practice. By segmenting the sample into neighbourhood typologies, the sample size of each model is reduced; with some clusters having very small samples. There is a high risk of over-specification given the number of variables included in the model, especially among the smaller clusters. This limits the robustness of findings about individual variables; which may depend on the specific model conditions (including interdependence of variables) rather than reflecting actual relationships (Alonso 1968). Nevertheless, the results from models IV to IX suggest that the relationships observed in one part of a city, for example downtown areas, do not necessarily reflect those in another, such as outer suburban development. This finding is consistent with recent empirical research that shows there is a non-linear association in space, between the built environment and public transport use (Ding et al. 2021).

5. Conclusion

This study aimed to understand the spatially varying and synergistic effects of the built environment on bus ridership at the stop level. We segmented our sample according to distinctive patterns of the built environment in bus catchments. In this way, we reduced the spatial heterogeneity of each ridership model. The typologies also allowed us to examine the overall effect of neighbourhood typology on ridership.

This study tested four approaches for exploring the relationship between the built environment and bus use in a city with diverse neighbourhood characteristics. Conventional approaches to analysis involve specifying multivariate models with many individual and often interdependent built environment variables. To simplify the analysis process and reduce multicollinearity, we instead proposed an approach that involves classifying bus station areas according to built environment characteristics, its ‘neighbourhood typology’, and using this to predict demand. We compare this approach to the conventional approach, for a sample of over 9,000 bus stops in Greater Melbourne. A fourth method was also developed to explore variation in the demand predictors between the neighbourhood typologies (Table 4).

We found that neighbourhood typology was significant for predicting bus use when considering both the supply and demographic attributes of the station area and treating the Fringe suburban typology as the reference. The model which used ‘typology’ as the only measure of the built environment explained almost as much variance in bus ridership as the model with all individual built environment variables. Depending on whether or not we include individual built environment attributes in the model, we find different results for the relative importance of neighbourhood typology for ridership. Results suggest there may be a latent demand for public transport in certain areas, which is not supported by the built environment. These nuances suggest that the overall character of the station environment is an important explanatory variable.

Results from a series of models developed for each neighbourhood typology suggest that the relationships observed in one part of a city do not necessarily reflect those in another, where the built environment and sociodemographic characteristics vary significantly. However, in situations where it may be of interest to identify appropriate interventions for a particular site, it may be appropriate to collect and examine data for similar locations only. Nevertheless, predicting demand using neighbourhood typology instead of individual variables is a simple approach that accounts for spatial variation while also yielding reliable results for all parts of a city. The typological approach therefore offers a simple method to segment demand analysis for public transport, including bus. The approach developed in this paper can be adapted by agencies to develop prediction models based on typologies that reflect those aspects of the built environment, or even policy, which can be influenced by interventions.

In this aggregate study, we did not have access to the kind of individual data that might have enabled us to separate the effects of self-selection from neighbourhood influence on travel behaviour. We also attempted to develop instrumental variables for public transport supply. However, multi-collinearity pervaded our models; and made our instrumented variable inappropriate. Thus, the impact of the built environment may be biased or masked. More work is needed to develop appropriate methods to account for this endogeneity issue in future research.

6. References

Alonso, W 1968, 'Predicting Best with Imperfect Data', *Journal of the American Institute of Planners.*, vol. 34, no. 4, pp. 248-55.

Aston, L, Currie, G, Delbosc, A, Fournier, N, Kamruzzaman, M & Teller, D 2020a, 'Melbourne built environment and transit use empirical dataset', electronic data set, *Monash University*, doi: https://doi.org/10.26180/5d9994f4704ea

Aston, L, Currie, G, Kamruzzaman, M & Delbosc, A 2021, 'Methodologies for empirical research on the link between the built environment and transit use', *Handbook of Public Transport Research*, Edward Elgar Publishing, Melbourne, Australia.

Aston, L, Currie, G, Kamruzzaman, M, Delbosc, A, Fournier, N & Teller, D 2020b, 'Addressing transit mode location bias in built environment-transit mode use research', *Journal of Transport Geography*, vol. 87, p. 102786.

Beck, MJ, Hensher, DA & Wei, E 2020, 'Slowly coming out of COVID-19 restrictions in Australia: Implications for working from home and commuting trips by car and public transport', *Journal of Transport Geography*, vol. 88, p. 102846.

Berrebi, S & Watkins, K 2020, 'Who's Ditching the Bus?', *Transportation Research Part A*, vol. 136, pp. 21-34.

Blitz, A & Lanzendorf, M 2020, 'Mobility design as a means of promoting non-motorised travel behaviour? A literature review of concepts and findings on design functions', *Journal of Transport Geography*, vol. 87. p. 102778.

Boulange, C, Gunn, L, Giles-Corti, B, Mavoa, S, Pettit, C & Badland, H 2017, 'Examining associations between urban design attributes and transport mode choice for walking, cycling, public transport and private motor vehicle trips', *Journal of Transport & Health*, vol. 6, pp. 155-66.

Cervero, R 2007, 'Transit-Oriented Development's Ridership Bonus: A Product of Self-Selection and Public Policies', *Environment and Planning A*, vol. 39, no. 9, pp. 2068-85.

Currie, G 2006, 'Bus transit oriented development—strengths and challenges relative to rail', *Journal of Public Transportation*, vol. 9, no. 4, p. 1.

Currie, G & Delbosc, A 2011, 'Understanding bus rapid transit route ridership drivers: An empirical study of Australian BRT systems', *Transport Policy*, vol. 18, no. 5, p. 755.

De Gruyter, C, Saghapour, T, Ma, L & Dodson, J 2020, 'How does the built environment affect transit use by train, tram and bus?', *Journal of Transport and Land Use*, vol. 13, no. 1, pp. 625-50.

de Haas, M, Faber, R & Hamersma, M 2020, 'How COVID-19 and the Dutch ‘intelligent lockdown’ change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands', *Transportation Research Interdisciplinary Perspectives*, vol. 6, p. 100150.

Diab, E, Kasraian, D, Miller, EJ & Shalaby, A 2020, 'The rise and fall of transit ridership across Canada: Understanding the determinants', *Transport Policy*, vol. 96, pp. 101-12.

Dill, J, Mohr, C & Ma, L 2014, 'How can psychological theory help cities increase walking and bicycling?', *Journal of the American Planning Association*, vol. 80, no. 1, pp. 36-51.

Ding, C, Cao, X, Yu, B & Ju, Y 2021, 'Non-linear associations between zonal built environment attributes and transit commuting mode choice accounting for spatial heterogeneity', *Transportation Research Part A: Policy and Practice*, vol. 148, pp. 22-35.

Ewing, R & Cervero, R 2010, 'Travel and the built environment: A meta-analysis', *Journal of the American Planning Association*, vol. 76, no. 3, pp. 265-94.

Flyvbjerg, B, Skamris Holm, MK & Buhl, SL 2005, 'How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation', *Journal of the American Planning Association*, vol. 71, no. 2, pp. 131-46.

Fuller, RJ & Crawford, RH 2011, 'Impact of past and future residential housing development patterns on energy demand and related emissions', *Journal of Housing and the Built Environment*, vol. 26, pp. 165–83.

Hair, JF, Black, WC, Babin, BJ & Anderson, RE 2014, *Multivariate data analysis*, Pearson Education Limited, Harlow, Essex.

Hall, JD, Palsson, C & Price, J 2018, 'Is Uber a substitute or complement for public transit?', *Journal of Urban Economics*, vol. 108, pp. 36-50.

Higgins, C & Kanaroglou, P 2016, 'A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region', *Journal of Transport Geography*, vol. 52, pp. 61-72.

Holz-Rau, C & Scheiner, J 2019, 'Land-use and transport planning – A field of complex cause-impact relationships. Thoughts on transport growth, greenhouse gas emissions and the built environment', *Transport Policy*, vol. 74, pp. 127-37.

Ibraeva, A, Correia, GHdA, Silva, C & Antunes, AP 2020, 'Transit-oriented development: A review of research achievements and challenges', *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 110-30.

Jeffrey, D, Boulangé, C, Giles-Corti, B, Washington, S & Gunn, L 2019, 'Using walkability measures to identify train stations with the potential to become transit oriented developments located in walkable neighbourhoods', *Journal of Transport Geography*, vol. 76, pp. 221-31.

Kamruzzaman, M, Baker, D, Washington, S & Turrell, G 2014, 'Advance transit oriented development typology: case study in Brisbane, Australia', *Journal of Transport Geography*, vol. 34, pp. 54-70.

Maier, MJ 2019, *REdaS: R: Einführung durch angewandte Statistik*, Pearson Studies, Munich.

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*, vol. 32, no. 1, pp. 33-46.

Pemberton, S 2020, 'Optimising Melbourne's bus routes for real-life travel patterns', *Case Studies on Transport Policy*, vol. 8, no. 3, pp. 1038-52.

Revelle, W 2018a, 'Exploratory Factor Analysis Using MinRes (Minimum Residual) As Well As EFA By Principal Axis, Weighted Least Squares Or Maximum Likelihood', *RDocumentation*, viewed 11 October 2021 , https://www.rdocumentation.org/packages/psych/versions/1.8.12/topics/fa

Rodríguez, DA & Kang, C-D 2020, 'A typology of the built environment around rail stops in the global transit-oriented city of Seoul, Korea', *Cities*, vol. 100, p. 102663.

Stojanovski, T 2018, 'How density, diversity, land use and neighbourhood typology influences bus mobility in the Swedish city of Karlstad: Mixing spatial analytic and typo-morphological approaches to assess the indirect effect of urban form on travel', *Journal of Transport and Land Use*, vol. 11, no. 1 pp. 769 - 789.

Taylor, BD, Miller, D, Iseki, H & Fink, C 2009, 'Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas', *Transportation Research Part A: Policy and Practice*, vol. 43, no. 1, pp. 60-77.

Tirachini, A & Cats, O 2020, 'COVID-19 and public transportation: current assessment, prospects and research needs', *Journal of Public Transportation*, vol. 22, no. 1, p. 1.

Vecchio, G & Martens, K 2021, 'Accessibility and the Capabilities Approach: a review of the literature and proposal for conceptual advancements', *Transport Reviews*, vol. 41, no. 1, pp. 1-22.

Voulgaris, CT 2019, 'Crystal Balls and Black Boxes: What Makes a Good Forecast?', *Journal of Planning Literature*, vol. 34, no. 3, pp. 286-99.

Voulgaris, CT, Taylor, BD, Blumenberg, E, Brown, A & Ralph, K 2017, 'Synergistic neighborhood relationships with travel behavior: An analysis of travel in 30,000 US neighborhoods', *Journal of Transport and Land Use*, vol. 10, no. 1, pp. 437-61.