

The value of accessibility in residential property

Renan Grace¹, Meead Saberi²

¹KPMG Australia

²University of New South Wales

Email for correspondence: rgrace@kpmg.com.au

Abstract

The impact of transport accessibility on land and property values and development densities is well known and widely acknowledged among practitioners and researchers in land use and transport planning. Quantitative methods to examine and predict these impacts are becoming increasingly relevant in the Australian context. This paper uses a hedonic regression to explore the impacts of car, public transport and walk accessibility on residential property sales values in Melbourne, Australia. The analysis controls for a number of other known drivers of property values. The analysis also accounts for spatial autocorrelation. A spatial autoregressive maximum likelihood model for house sale prices between 2012 and 2016 is presented. Accessibility metrics are measured on a scale of 0 to 1. Each 0.1 increase in car, public transport and walk accessibility is associated with a 6.8%, 0.4% and 0.9% increase in house sale values respectively.

1. Introduction

The impact of transport accessibility on land use is well understood and widely acknowledged among transport and land use planners in government, academia and consulting. The concept of accessibility and its impact on land use was first popularised in the seminal 1959 paper “How accessibility shapes land use” by American planner Walter G. Hansen (Hansen, 1959).

Despite this, most project appraisals for major infrastructure projects assume no change in land use between Base and Project Case. For major projects, this is a limitation which has the potential to bias evidence used to select and justify projects.

The “city-shaping” influence of major infrastructure projects is growing in recognition and is beginning to influence investment cases. Infrastructure Australia now requires proponents to consider tolling or value capture mechanisms in business cases for initiatives and projects to be considered for the Infrastructure Priority List (Infrastructure Australia, 2017). This recognises the impact that transport infrastructure has on land and property values.

In addition, the existence of “wider economic benefits” (WEBs) of transport projects is gaining recognition in Australia. The Australian Transport Assessment and Planning (ATAP) Guidelines are in the process of preparing guidelines for the estimation of WEBs in Australian cities. Some categories of WEB require the estimation of land use changes induced by a transport infrastructure project, including “WB1: agglomeration economies” (cluster effects) and “WB2b: move to more productive jobs” (KPMG, 2017).

Land use changes can be estimated using a “land use transport interaction” (LUTI) model. The estimation of changes in land and property values are a key component of LUTI models. The UrbanSim framework is commonly used for LUTI modelling in the USA and Europe. In addition to the transport model, UrbanSim implementations commonly include hedonic price modules to represent property values, “relocation and transition” modules to represent changes in population and employment, “location choice” modules to represent consumers’ and businesses’ location decisions and a “developer activity” module to represent supply decisions made by developers (Waddell, 2011). The function of the hedonic pricing module within a LUTI model framework is to mediate changes in land use – attractive areas increase in value, which puts a ‘brake’ on additional development.

The objective of this work is to specify a residential hedonic pricing regression for assessing the impact of transport accessibility on land and property values in Melbourne, Australia. A hedonic pricing approach separates the various components that drive property values. It assumes that different attributes of a property (e.g. number of bedrooms, accessibility to jobs) have a cumulative effect on the property value. The work is intended to be able to be applied for value capture analysis and as a component for LUTI models.

The remainder of this paper is structured as follows:

- Literature Review
- Methodology
- Results
- Discussion
- Conclusion and next steps

2. Literature review

2.1 Accessibility and property values

Accessibility can be defined as the ‘potential of opportunities for interaction’ (Hansen, 1959). In this context, accessibility represents the extent to which a person, at a given place and time, has the ability to access opportunities that they want or need to access. For example, employment is a specific type of opportunity that many people seek access to. Other opportunities include health services, education, social, recreational and leisure activities. Individuals seek out, and derive utility from, accessibility to opportunities.

The common maxim in real estate “location, location, location” reflects the importance people assign to accessibility in forming preferences about where to purchase property. Well located properties are valuable in large part because they provide residents with high quality access to opportunities (e.g. jobs and services).

2.2 Cross-sectional and longitudinal analyses

A large body of research exists exploring the relationship between accessibility and the value of land and property. There is significant evidence that accessibility contributes positively to property values.

Existing studies can be divided into two broad categories:

- Cross-sectional analyses consider how property values correlate with accessibility at a single point in time. The ability to draw conclusions about the direction of causality is limited in cross-sectional analyses.

- Longitudinal analyses consider how changes in property values relate to changes in accessibility at a given location (for example due to a new public transport corridor) over two or more points in time. Longitudinal analyses provide a more robust basis for inferring the causal relationship between accessibility and property values.

While longitudinal analyses are theoretically preferable, lack of availability of accurate and consistent data for the performance of transport networks at a local level over time can limit the ability for researchers to conduct this type of analysis. As a result, cross-sectional analyses are more common. In the era of smart phones and “big data”, the availability of high quality time series transport network performance data is likely to increase in coming years. This may improve the capacity of researchers to undertake longitudinal analyses.

Evidence of a positive relationship between accessibility and property values exists for both cross-sectional (LUTI Mecone, 2016; Iacono & Levinson, 2017; Giuliano, et al., 2010; Martinez & Viegas, 2009; Armstrong & Rodriguez, 2006; Srour, et al., 2002) and longitudinal analyses (LUTI Mecone, 2016; Mulley, 2014; Boucq, 2007).

Iacono & Levinson (2017) found positive impacts of accessibility (using a cumulative opportunities measure) on property values for a cross-sectional analysis. However, this effect was not apparent when a longitudinal analysis was undertaken using the same data. It was postulated that this may be due to diminishing returns of marginal accessibility improvements in mature networks. The use of accessibility measures with a saturation function as proposed by Espada & Luk (2011) may be useful to control for this effect, as these measures explicitly account for diminishing marginal returns of improved accessibility.

2.3 Measuring accessibility

Geurs and van Wee (2004) identified four components that are theoretically important to be included in an accessibility measure:

- The *land-use component* refers to the spatial distribution and quality of opportunities.
- The *transportation component* refers to the disutility experienced in travelling from a given location to a relevant opportunity. Measures of disutility may include travel time, costs and perceived inconveniences like transfers.
- The *temporal component* refers to the availability of opportunities at different times of day, and the time available for people to participate in those opportunities.
- The *individual component* refers to the needs and preferences of different individuals. For example, how far a person is willing to travel, their car availability or their level of education and skill in assessing which employment opportunities are available to them.

Numerous accessibility metrics have been defined and measured for various purposes. Geurs & van Wee (2004) characterise four types of accessibility measures:

- *Infrastructure-based measures* consider the performance of a specific piece of infrastructure. For example, “level of service” and “average speed”.
- *Location-based measures* consider the array of opportunities available from a given location. For example, “number of jobs within 30 minutes”.
- *Person-based measures* consider the opportunities available to a given individual at a given time.
- *Utility-based measures* consider the economic benefit that people derive from access to opportunities.

Location-based measures are appropriate for considering the impact of accessibility on land and property values and development densities, as land use is fixed in location by its nature, and may be used by a variety of individuals.

Espada & Luk (2011) developed a location-based accessibility metric suitable for Melbourne known as the ARRB Accessibility Metric (AAM). The AAM is an example of a ‘cumulative opportunities’ measure. Cumulative opportunities measures consider the number of opportunities (potential destinations) available from a given origin, typically weighted by an impedance decay. The AAM is similar to other cumulative opportunities accessibility measures used in the literature (e.g. Anderson et al. 2013, Melo et al. 2017). The AAM meets all four of the criteria for a theoretically robust accessibility measure as defined by Geurs and van Wee (2004).

The AAM uses a ‘deterrence function’ to reflect that the value of an opportunity declines as the travel impedance between a given origin and that opportunity increases. For example, a job accessible within 30 minutes is ‘worth more’ than a job that is accessible in 60 minutes, but both jobs have opportunity value.

Another key feature of the AAM is its specification of a ‘saturation function’ to reflect diminishing marginal returns of additional opportunities. This reflects that the availability of additional opportunities has declining marginal value. The value of additional activities tends to zero as the number of opportunities approaches infinity. Espada & Luk (2011) calibrated the saturation curve using the Melbourne Integrated Transport model, a strategic four step model held by the Victorian Government.

The AAM specifies deterrence and saturation functions for a variety of accessibility measures, including for different opportunities (e.g. work, education, shopping and recreation) as well as for different modes (e.g. car, public transport, cycle, walk).

2.4 Quantitative methods

The most common quantitative method for estimating the impact of accessibility on property values is the Hedonic Price Method. The Hedonic Price Method assumes that the value of a good can be broken down into constituent components that each contribute to the total value. An advantage of the Hedonic Price Method is that it allows various components of a property (e.g. number of bedrooms, accessibility to jobs, etc.) to be assigned a separate value. The vast majority of studies considering the impact of accessibility on property value use a linear regression hedonic pricing specification. Other potential methods include multilevel models (e.g. random coefficient models) (Giuliano, et al., 2010), though these are less common in the literature.

2.5 Controlling for other drivers of property value

In order to estimate the impact of accessibility on property values using the Hedonic Price method, it is necessary to control for the impact of other known factors that influence property values.

The following factors non-accessibility-related factors are demonstrated in the literature to impact on property values:

- Structural features of the property (e.g. number of bedrooms, number of bathrooms, floor space, lot size) (LUTI Mecone, 2016; Iacono & Levinson, 2017; Giuliano, et al., 2010)
- Socio-economic characteristics of a neighbourhood (e.g. average income, mix of occupations, education levels, crime levels, perceived vitality) (Harris, 1999; De Nadai & Lepri, 2018)
- Whether a property is within the catchment area of high quality educational institutions (Collins & Kaplan, 2017)
- Proximity to coastline (Giuliano, et al., 2010)
- Adjacency to major roads (LUTI Mecone, 2016)

In addition, property values change over time, necessitating the use of time dummy variables which reflect the timing of property sales where the data is collected over a sufficiently long time period (Hansen, 2009).

2.6 Spatial autocorrelation

Spatial autocorrelation is when a functional relationship exists between values that are located nearby to each other in space. Several studies have demonstrated the existence of spatial autocorrelation in Ordinary Least Squares (OLS) hedonic price models which specify property values as an independent variable (Diao, 2015; Basu & Thibodeau, 1998). As a result, many existing studies that consider the impact of accessibility on property values attempt to control for spatial autocorrelation in some way (Diao, 2015; Mulley, 2014; Martinez & Viegas, 2009; Armstrong & Rodriguez, 2006).

The existence of spatial autocorrelation suggests that significant factors that influence property values are missing from the OLS model, and these factors are spatially correlated. Some potential examples of this are the style and era of architecture, the quality of streetscapes and the perceived “niceness”, prestige or history of a neighbourhood. While it is difficult to measure these factors objectively and accurately for incorporation into an OLS model, they can be accounted for indirectly by using a spatial autoregressive model.

The OLS model framework is not able to be adjusted to account for spatial autocorrelation. The inclusion of a spatially lagged dependent variable in OLS produces biased and inconsistent parameter estimates (Anselin, 1988). The presence of spatial residual autocorrelation results in inefficient parameter estimates. Therefore, the use of a maximum likelihood approach for spatial autoregressive model is required.

Scripts are available in the R statistical package library “spdep” (spatial dependency) to estimate spatial autoregressive models using a maximum likelihood approach (Bivand, et al., 2008).

3. Methodology

3.1 Model form

This study uses a cross-sectional hedonic pricing regression. While it is acknowledged that longitudinal analyses are theoretically preferable for inferring causation, high quality time-series data on transport network performance was not available for this study.

Initially, an OLS regression was fitted with the logarithm of sales price as the dependent variable, and a set of property characteristics as independent variables. The model form is shown in Equation 1.

$$\ln \mathbf{y} = \mathbf{X}\beta + \mathbf{e} \quad (1)$$

Where $\ln \mathbf{y}$ is a vector of the logarithm of sales prices (the dependent variable), \mathbf{X} is a matrix of regressors (the independent variables) and \mathbf{e} is a vector of error terms.

Separate regressions were undertaken for the houses and units sub-markets.

After fitting the OLS regressions, Moran's I tests were conducted to detect the presence of spatial autocorrelation. Lagrange Multiplier tests were then conducted to identify the type of spatial autocorrelation present, and therefore the most appropriate spatial autoregressive model specification (Anselin, 1988).

Two model specifications were considered to control for different types of spatial autocorrelation. The first is a "spatial lag" model as shown in Equation 2, which incorporates a spatially lagged dependent variable. The second is a "spatial error" model, as shown in Equation 3, which accounts for spatial autocorrelation in the model residuals (Bivand, et al., 2008). The models can be executed in the open-source R Statistical Package using the library "spdep".

$$\ln \mathbf{y} = \rho \mathbf{W} \ln \mathbf{y} + \mathbf{X}\beta + \mathbf{e} \quad (2)$$

$$\ln \mathbf{y} = \mathbf{X}\beta + \mathbf{u} \quad (3)$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \mathbf{e}$$

Where \mathbf{W} is a spatial weights matrix. For this analysis, spatial weights are assumed to be uniform (and sum to one) for all observations within 400 metres of each dependent variable record for houses, and 200 metres for units.

The independent variables in the matrix \mathbf{X} can be separated into four categories as shown below:

- Time variables
- Structural variables
- Accessibility variables
- Neighbourhood variables

The remainder of this Section describes the variables used in the analysis.

3.2 Sales data

The residential property sale price data was extracted from CoreLogic's RP Data Property database. Permission was sought from and granted by CoreLogic to use the data for this analysis. Transactions within the Greater Melbourne area for sales that occurred between 2011 and 2017 were included. This database includes 528,993 records.

Spatial coordinates were obtained by geocoding the address strings in the sales database. A custom geocoder was prepared which matched the address strings to the VicMap Address

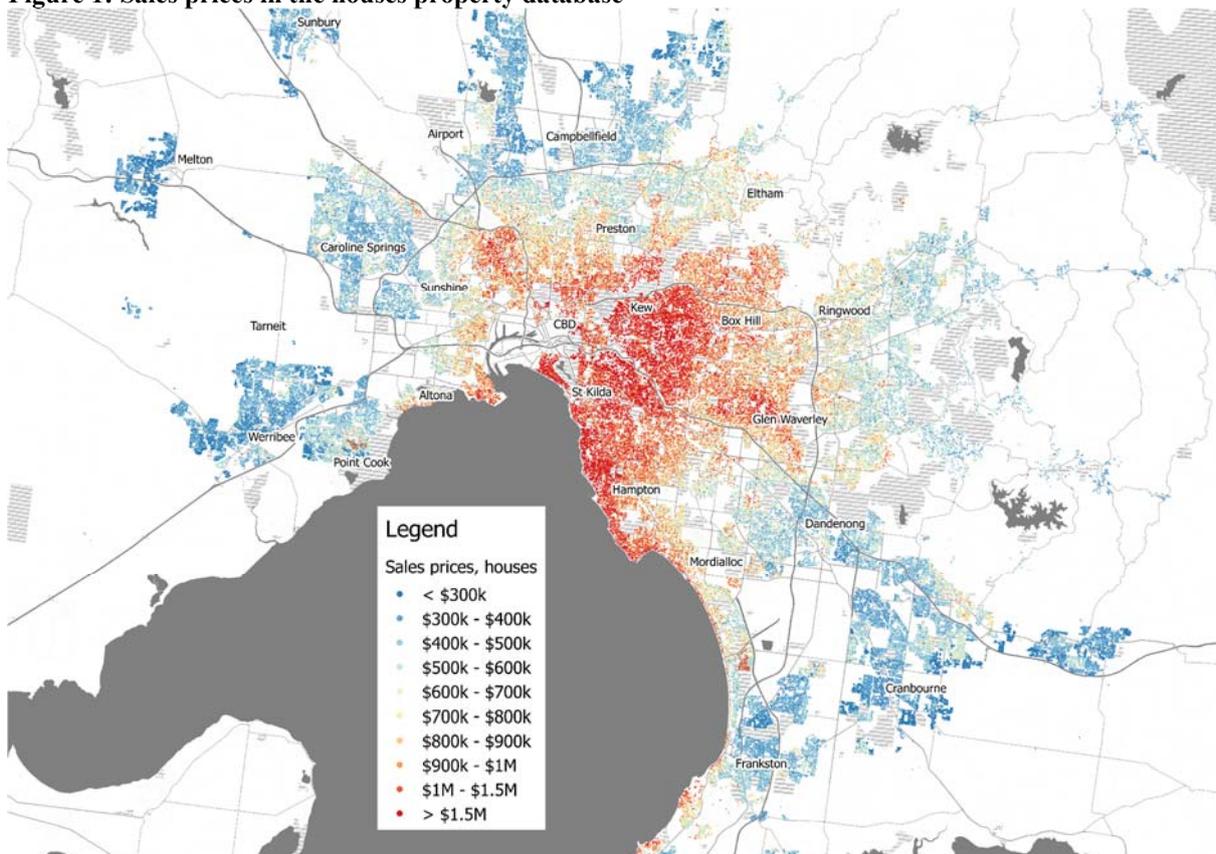
database obtained from the data.vic.gov.au website. Addresses that could not be obtained from the VicMap Address database were geocoded using the Google Maps API. Records with address strings that were unable to be geocoded by either method were discarded. 19,143 records (3.6%) were discarded, leaving 509,580 records remaining after geocoding.

In addition to the above, the dataset was filtered for sales prices between \$100k and \$10M, for one to eight bedrooms, for one to five bathrooms, that sold within the calendar years 2012 through 2016, and with locations inside the Melbourne Greater Capital City Statistical Area (GCCSA) as defined by the Australian Bureau of Statistics for the 2011 Census. In addition, houses with lot sizes less than 150 m² or more than 1500 m² (or no lot size recorded) and units with more than four bedrooms were excluded from the analysis.

The database was divided into two – one with records designated as “House” by CoreLogic and the other with records designated as “Unit”. Some records were designated as houses, but had the character ‘/’ in the address string (suggesting multiple dwellings on a parcel). These were assumed to be erroneously coded and were moved from the houses to the units database.

After filtering, the final houses database had 228,824 records and the final units database had 129,421 records. A representation of the houses database is shown in Figure 1.

Figure 1: Sales prices in the houses property database



Source: RP Data Property database, Corelogic

3.3 Time variables

A set of dummy variables was generated to represent the year of sale. The reference category was designated as sales from year 2016. The dummy variables for the remaining years were named y2012, y2013, y2014 and y2015.

3.4 Structural variables

Dummy variables were generated to account for the number of bedrooms for each sold property. The reference category was designated as three bedrooms for houses and two bedrooms for units. The dummy variables for the remaining categories were named bed12 (one and two bedroom houses), bed4 (four bedroom houses) and bed5plus (houses with five or more bedrooms). For units, bed1, bed2, bed3 and bed4plus were used.

A dummy variable was generated for houses and units with more than one bathroom. The dummy variable was named extra_baths.

The logarithm of lot size for houses was included, with the variable named ln(lotsize). No such variable was included for units.

3.5 Accessibility variables

The following accessibility variables were generated:

- Car accessibility to work car_acc
- Public transport accessibility to work pt_acc
- Walk accessibility to amenities walk_acc

The car accessibility, public transport accessibility and walk accessibility variables were derived using the AAM methodology (Espada & Luk, 2011). The accessibility indices are calculated using Equation 4. The AAM yields accessibility indices which are a continuous variable between 0.0 and 1.0. The accessibility outcomes are shown in Figures 2, 3 and 4.

$$A_i = s \left(\sum_j d(C_{ij})X_j \right) \quad (4)$$

Where s is the saturation function, d is the deterrence function, C_{ij} is the transport impedance between zones i and j , and X_j is the number of opportunities available at zone j .

The S-shaped deterrence function is calculated using Equation 5.

$$d(C) = \frac{e^{-\beta(C-\alpha)}}{1 + [e^{-\beta(C-\alpha)}]^\lambda} \quad (5)$$

$$\lambda = \frac{1 + e^{\alpha\beta}}{e^{\alpha\beta}}$$

Where α and β are parameters. The assumed values are sourced from Espada & Luk (2011) and are shown in Table 1.

Table 1: Deterrence function parameters

Measure	α	β
Car to work	25.5	0.14
Public transport to work	46.0	0.07
Walk to shopping and recreation	11.4	0.58

Congested car travel times were sourced from the Google Maps API “Distance Matrix” for a departure time of 8am. Predicted travel times for Tuesday 2nd May 2017 were used (collected in advance of that date). The date was selected as it falls outside of school holidays and is not in a week with any public holidays. In addition, May is usually considered as a “typical” month for weekday traffic congestion in Melbourne. Travel times predicted in advance were used to ensure that the values were not affected by any one-off events (eg. incidents, roadworks).

Travel times were collected for a 281 by 281 zone system, where the zone system reflects the 2011 Statistical Area 2 (SA2) boundaries as defined by the Australian Bureau of Statistics. The geographic centroids of each SA2 were used as origin and destination coordinates.

Public transport travel times were sourced from the KPMG’s Melbourne Activity and Agent Based Model (MABM). Travel times include walk, wait and in-vehicle times, and are averaged for AM departure times of 7:50, 7:55, 8:00, 8:05 and 8:10. Services are assumed to run as scheduled, with no delays or cancellations. A custom zone system was defined for public transport origins and destinations, with a 1km grid along major public transport corridors and a 5km grid elsewhere.

Costs are measured in generalised minutes. A value of time of \$16 per hour is assumed to convert monetary costs to generalised minutes as per Espada & Luk (2011). Parking and tolls are included for car trips and fares are included for public transport. No monetary costs are incurred for walking trips. All relevant monetary costs were sourced from the MABM. Half of daily parking costs were assumed, as accessibility estimates are calculated for a one-way trip.

Walk travel times were estimated using the pandana open-source software. Melbourne’s entire pedestrian network as defined on OpenStreetMap was converted to a network graph, allowing walk accessibility to be estimated at a fine resolution. Walk speed was assumed to be 4.3 km/h as specified in Espada & Luk (2011).

The saturation function is calculated as follows:

$$s = \frac{1 - e^{-\kappa \sum_j a(c_{ij})x_j}}{1 + e^{-\kappa \sum_j a(c_{ij})x_j}} \quad (6)$$

Where κ is a parameter relating to the opportunity type.

For car and public transport accessibility to work, opportunities are defined as jobs according to the Victoria in Future estimates for 2015 at a travel zone level (approximately 3,500 zones in the Melbourne GCCSA). A κ value of 10^{-5} is assumed as per Espada & Luk (2011).

For walk accessibility to amenities, opportunities are defined as any “Google Place” defined as “supermarket”, “bar”, “restaurant” or “sporting facility”. Google Places were sourced from the Google Maps Public API. A custom κ value of 0.04 is assumed which was calibrated by the author to provide an intuitive measure, given the custom nature of the opportunity measure.

Accessibility was assumed not to vary by year, due to lack of data regarding the change in performance of the transport network between 2011 and 2016.

Figure 2: Accessibility to jobs by car, 8am (*car_acc*)

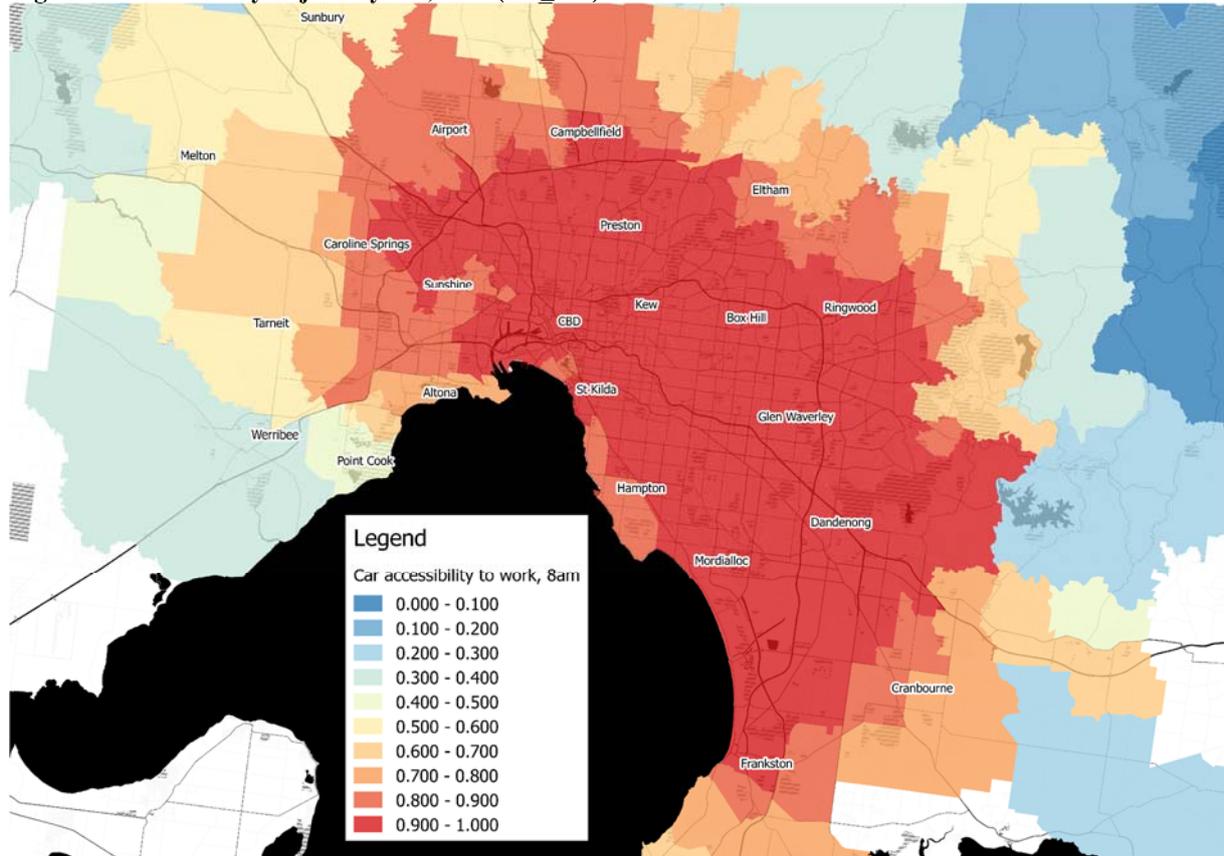


Figure 3: Accessibility to jobs by public transport, 8am (*pt_acc*)

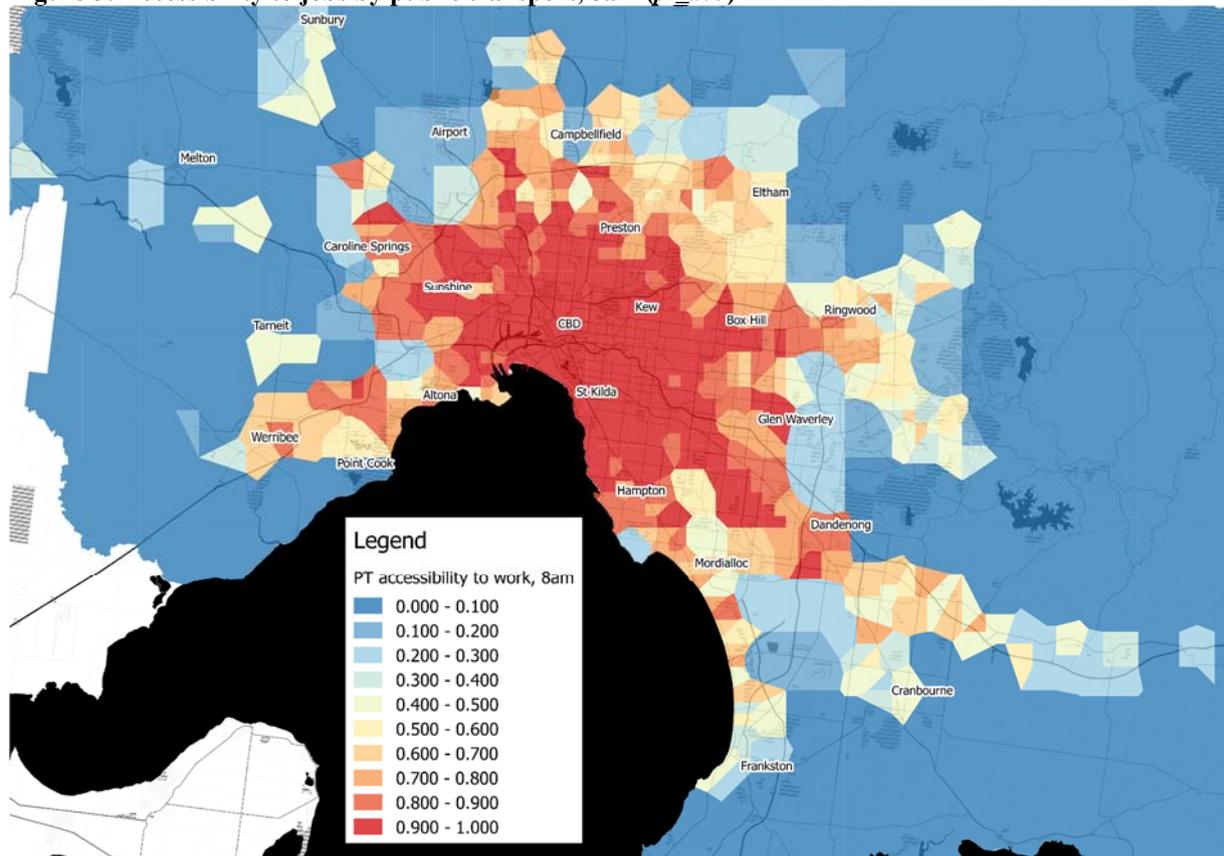
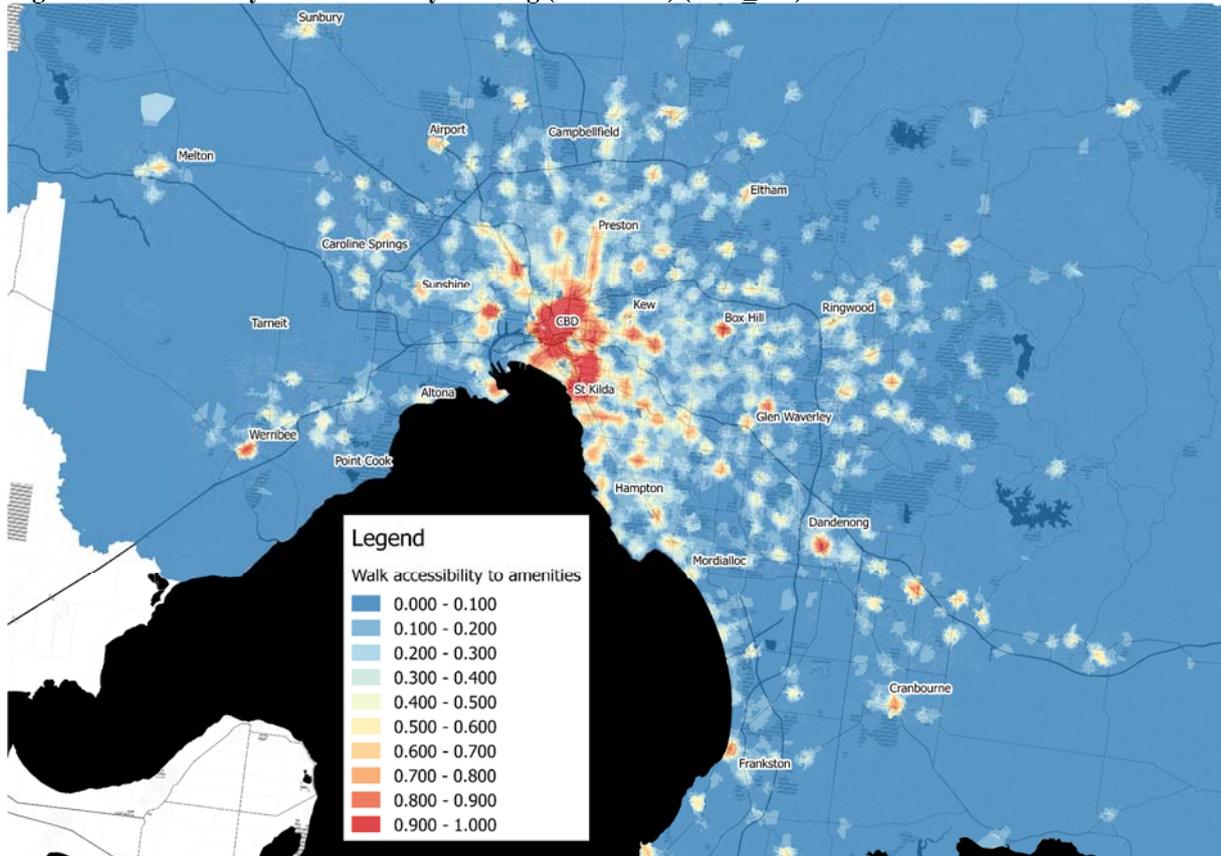


Figure 4: Accessibility to amenities by walking (4.3 km/hr) (*walk_acc*)



3.6 Neighbourhood variables

Four neighbourhood variables were defined to control for other known drivers of property values. The following variables were included:

- Socio-economic character of neighbourhood *irsad_q1, irsad_q2, irsad_q4*
- Within top public high school catchment *top25_highschool*
- Proximity to coastline *coast800*
- Adjacent to arterial road *arterial50*

The ABS 2011 Socio-Economic Indexes for Areas (SEIFA) Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD) was used to control for the socio-economic character of each neighbourhood (at the Statistical Area 2 level). Dummy variables were generated for each IRSAD quartile, with quartile 4 representing the most advantaged and least disadvantaged neighbourhoods and quartile 1 representing the least advantaged and most disadvantaged neighbourhoods. The reference variable is IRSAD quartile 3, making the dummy variable categories *irsad_q1, irsad_q2* and *irsad_q4*.

The location of all Government secondary schools in Victoria was sourced from data.vic.gov.au. Each sales record was assigned to the nearest Government high school by straight line distance. This is an approximate measure of which secondary school catchment a property lies within. The top 25 Government secondary schools (excluding selective schools) were identified using the “Overall Rating” from the “Better Education” website, which considers the level of academic achievement. Properties falling within the catchment of a top 25 Government secondary school (approximately the top ten percent of Victorian secondary

schools) were assigned a value of one for the dummy variable *top25_highschool*, with other properties assigned a value of zero for that variable.

A dummy variable *coast800* was specified with a value of one if the address point coordinate of a property falls within 800m straight line distance from the coast, and zero otherwise.

A dummy variable *arterial50* was specified with a value of one if the address point coordinate of a property falls within 50m of an arterial road. Arterial roads were defined as roads with a CLASSCODE of one, two or three (which does not include freeways) in the TR_ROAD dataset available on data.vic.gov.au. In practice, this generally means the property is directly adjacent to a major arterial road.

4. Results

4.1 OLS model results

The results for the OLS model specifications are shown in Table 4 for houses. The equivalent results for units are not shown due to their limited generalisability as discussed in the remainder of this section.

4.2 OLS statistical assumption testing

The OLS models were tested for the degree to which they meet the statistical assumptions of linearity of model, heteroscedasticity, normality of residuals, outliers and influential cases and multicollinearity. Statistical assumptions are met to a reasonable degree for the OLS models, however the houses OLS violates the assumption of homoscedasticity and the units OLS violates the assumptions of homoscedasticity and normality of residuals.

Violation of assumptions does not necessarily invalidate drawing of conclusions from the sample, but may affect the ability to generalise findings.

4.3 Spatial autocorrelation

Both the houses and units OLS models appeared to exhibit spatial autocorrelation of residuals. Moran's I tests were undertaken, which confirm the presence of autocorrelation in model residuals for both the houses and units OLS at a $p < 0.001$ level of confidence.

Lagrange multiplier tests were undertaken to determine which spatial autoregressive model specification is most appropriate. The results are shown in Table 2, with the test statistic clearly indicating that the spatial error model is the most appropriate specification for both houses and units.

Table 2: Lagrange multiplier diagnostics

Model	Test statistic	df	p-value
Houses, spatial lag	96,886	1	<2.2e-16 ***
Houses, spatial error	1,915,061	1	<2.2e-16 ***
Units, spatial lag	4,689	1	<2.2e-16 ***
Units, spatial error	56,369	1	<2.2e-16 ***

*Significance codes: 0.001 '***', 0.01 '**', 0.05 '*', 0.1 '.'*

Spatial error models were run for both houses and units. The results are shown in Table 4 for houses and not shown for units. In both cases, the spatial error model produces a substantially better fit to observed data than the OLS, as evidenced by lower Akaike Information Criterion (AIC) statistics. While both fits are improved, the fit is improved to a substantially greater degree for houses than units. In addition, a noticeable reduction in spatial autocorrelation is visibly apparent in the residuals. Spatial autocorrelation is reduced to a greater extent for houses than units.

The houses spatial error model also exhibit a substantially reduced level of heteroscedasticity relative to the houses OLS model. This is demonstrated by a reduction in the studentised Breusch-Pagan test statistics as shown in Table 3. Heteroscedasticity is also reduced by the spatial error model for units, but to a lesser degree.

Table 3: Studentised Breusch-Pagan diagnostics

Model	Test statistic	df	p-value
OLS, houses	8,348.28	18	<2.2e-16 ***
Spatial error, houses	686.66	18	<2.2e-16 ***
OLS, units	1,699.07	17	<2.2e-16 ***
Spatial error, units	1,298.67	17	<2.2e-16 ***

*Significance codes: 0.001 '***', 0.01 '**', 0.05 '*', 0.1 '.'*

Given the reduction in spatial autocorrelation and heteroscedasticity, the spatial error model for houses is deemed statistically robust and suitable for generalising findings. The other three models (OLS houses, OLS units, spatial error units) have statistical properties which limit the generalisability of findings.

Table 4: Model fits for OLS and spatial error models, houses

	OLS, houses				Spatial error, houses			
	Estimate	Std. Error	t-value	Pr(> t)	Estimate	Std. Error	z-value	Pr(> z)
Intercept	10.509	0.0127	824.82	<2e-16 ***	10.946	0.0323	338.71	<2e-16 ***
Time variables								
<i>y2012</i>	-0.3466	0.0024	-147.09	<2e-16 ***	-0.3641	0.0016	-231.50	<2e-16 ***
<i>y2013</i>	-0.3027	0.0022	-136.44	<2e-16 ***	-0.3136	0.0015	-211.72	<2e-16 ***
<i>y2014</i>	-0.2357	0.0021	-110.30	<2e-16 ***	-0.2354	0.0014	-164.94	<2e-16 ***
<i>y2015</i>	-0.1181	0.0021	-56.96	<2e-16 ***	-0.1106	0.0014	-80.11	<2e-16 ***
Structural variables								
<i>bed12</i>	0.0301	0.0030	10.18	<2e-16 ***	-0.0618	0.0020	-30.63	<2e-16 ***
<i>bed4</i>	0.0578	0.0017	34.41	<2e-16 ***	0.0669	0.0012	57.45	<2e-16 ***
<i>bed5plus</i>	0.1968	0.0031	62.99	<2e-16 ***	0.1406	0.0021	66.78	<2e-16 ***
<i>ln(lotsize)</i>	0.3035	0.0019	161.66	<2e-16 ***	0.2859	0.0017	171.88	<2e-16 ***
<i>extra_baths</i>	0.1423	0.0018	77.91	<2e-16 ***	0.1089	0.0013	83.67	<2e-16 ***
Accessibility variables								
<i>car_acc</i>	0.6350	0.0046	139.31	<2e-16 ***	0.6762	0.0367	18.42	<2e-16 ***
<i>pt_acc</i>	0.4699	0.0028	166.89	<2e-16 ***	0.0406	0.0086	4.69	2.69e-6 ***
<i>walk_acc</i>	0.7475	0.0049	153.30	<2e-16 ***	0.0886	0.0114	7.76	8.66e-15 ***
Neighbourhood variables								
<i>irsad_q1</i>	-0.2465	0.0020	-122.55	<2e-16 ***	-0.1071	0.0082	-13.09	<2e-16 ***
<i>irsad_q2</i>	-0.0726	0.0021	-35.05	<2e-16 ***	-0.0417	0.0060	-6.90	5.35e-12 ***
<i>irsad_q4</i>	0.3091	0.0022	141.89	<2e-16 ***	0.0451	0.0066	6.79	1.12e-11 ***
<i>top25_highschool</i>	0.1862	0.0023	81.88	<2e-16 ***	0.0223	0.0064	3.49	4.80e-4 ***
<i>coast800</i>	0.3934	0.0036	108.09	<2e-16 ***	0.0134	0.0082	1.63	0.1025
<i>arterial50</i>	-0.0328	0.0026	-12.45	<2e-16 ***	-0.0585	0.0019	-31.18	<2e-16 ***
Residual std. error	0.3323 on 228,805 degrees of freedom							
Adjusted R ²	0.6961							
Pseudo-R ²	0.8593							
F-statistic	2.91e+04 on 18 and 228,805 DF, p-value: < 2.2e-16							
Wald statistic	2,052,700, p-value: < 2.22e-16							
AIC	145,220 -30,922							

Significance codes: 0.001 '***', 0.01 '**', 0.05 '*', 0.1 '.'

5. Discussion

This discussion draws on the results from the houses spatial error (cross-sectional) model, which is deemed statistically robust. The results show that all three accessibility measures (walk to amenities `walk_acc`, public transport to jobs `pt_acc` and car to jobs `car_acc`) are positively associated with house sale prices. This finding is generally consistent with the literature. Car accessibility has the strongest influence, with a semi-elasticity of 0.68 (i.e. each 0.1 unit increase in car accessibility raises house sale prices by approximately 6.8%). Walk and public transport accessibility semi-elasticities are substantially lower, at 0.09 and 0.04 respectively.

This finding suggests that Melbourne homebuyers have a lower willingness to pay for walk and public transport accessibility than car accessibility. This is consistent with the car-dependent nature of Melbourne – car ownership is high in Melbourne, and Melburnians undertake 72% of trips and 82% of kilometres by private car (Victorian Government, 2013).

This study also highlights the importance of appropriately controlling for spatial autocorrelation in hedonic pricing analysis for property sales. Using an OLS model specification produced significantly different magnitudes of parameter estimates, particularly for measures with a spatial element (i.e. accessibility and neighbourhood variables). For the `bed12` dummy variable, the sign of the parameter estimate was inconsistent with theory, with fewer bedrooms having a positive influence on house sales values (the `bed12` dummy variable represents one and two bedroom houses compared to the reference category of three bedroom houses). This effect was (appropriately) no longer apparent after controlling for spatial autocorrelation.

For both OLS and spatial error model specifications, model performance was substantially better for house sales than unit sales, both in terms of predictive power and statistical assumptions. A potential explanation for this is that attributes that are not reflected in the dataset are important drivers of unit prices. For example, floor space, views and features of the apartment building itself (attractiveness, availability of services like a concierge and facilities like shared spaces, pools or gyms) are not included in the analysis. In addition, the units database also includes townhouses, which may be considered a separate sub-market, but are unable to be distinguished. These issues limit the generalisability of the units models.

6. Conclusions and next steps

This study presents a statistically robust hedonic price model that may be used to inform value capture and LUTI modelling studies in Melbourne using trip-based measures of accessibility which are consistent with most strategic transport models used in Australia. It is recommended that only the results of the houses spatial error model be generalised. The houses spatial error model produced semi-elasticities of car accessibility to jobs, public transport accessibility to jobs and walk accessibility to amenities of 0.68, 0.04 and 0.09 respectively.

Further work may produce a longitudinal analysis to estimate how changes in accessibility impact changes in property values (for example using repeat sales measures). This work would require high quality time series transport network performance data, and would need to control for changes in the quality of properties (eg. renovations) and neighbourhoods (eg. parks and gardens, schools, town centres) in addition to changes in accessibility.

Further work may also include testing of alternative specifications of accessibility metrics and control variables, or specifying additional control variables.

7. Acknowledgements

We would like to acknowledge and thank CoreLogic for granting permission to use the RP Data Property product transaction data for this study.

8. References

- Anderson, P., Levinson, D. & Parthasarathi, P., 2013. Accessibility Futures. *Transactions in GIS*, 17(5), pp. 683-705.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Santa Barbara: Department of Geography and Economics, University of California.
- Armstrong, R. & Rodriguez, D., 2006. An Evaluation of the Accessibility Benefits of Commuter Rail in Eastern Massachusetts using Spatial Hedonic Price Functions. *Transportation*, 33(1), pp. 21-43.
- Basu, S. & Thibodeau, T. G., 1998. Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 17(1), pp. 61-85.
- Bivand, R. S., Pebesma, E. J. & Gomez-Rubio, V., 2008. *Applied Spatial Data Analysis with R*. New York: Springer.
- Boucq, E., 2007. *The effects of accessibility gains on residential property values in urban area: the example of T2 tramway in Hauts-de-Seine department, France*. Leiden, The Netherlands, The European Transport Conference (ETC) held 17-19 October 2007.
- Collins, C. A. & Kaplan, E. K., 2017. Capitalization of school quality in housing prices: Evidence from boundary changes in Shelby County, Tennessee. *American Economic Review: Papers & Proceedings*, 107(5), pp. 628-632.
- De Nadai, M. & Lepri, B., 2018. *The economic value of neighbourhoods: Predicting real estate prices from the urban environment*. Turin, Italy, The IEEE DSAA, 2018 held 1-4 October 2018.
- Diao, M., 2015. Selectivity, spatial autocorrelation and the valuation of transit accessibility. *Urban Studies*, 52(1), pp. 159-177.
- Espada, I. & Luk, J., 2011. Development of an accessibility metric and its application to Melbourne. *Road & Transport Research*, 20(3), pp. 66-77.
- Geurs, K. & van Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, Volume 12, pp. 127-140.
- Giuliano, G., Gordon, P., Pan, Q. & Park, J., 2010. Accessibility and Residential Land Values: Some Tests with New Measures. *Urban Studies*, 47(10), pp. 3103-3130.
- Hansen, J., 2009. Australian house prices: A comparison of hedonic and repeat-sales measures. *The Economic Record*, 85(269), pp. 132-145.
- Hansen, W. G., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), pp. 73-76.
- Harris, D. R., 1999. "Property Values Drop When Blacks Move in, Because...": Racial and Socioeconomic Determinants of Neighbourhood Desirability. *American Sociological Review*, 64(3), pp. 461-479.
- Iacono, M. & Levinson, D., 2017. Accessibility dynamics and location premia: Do land values follow accessibility changes?. *Urban Studies*, 54(2), pp. 364-381.
- Infrastructure Australia, 2017. *Assessment framework for initiatives and projects to be included in the Infrastructure Priority List*, Sydney: Commonwealth of Australia.
- KPMG, 2017. *Measuring WEBS in Australian cities*, Melbourne: KPMG.

- LUTI Mecone, 2016. *Transit and Urban Renewal Value Creation*, Sydney: LUTI Consulting.
- Martinez, M. L. & Viegas, J. M., 2009. Effects of Transportation Accessibility on Residential Property Values Hedonic Price Model in the Lisbon, Portugal, Metropolitan Area. *Transportation Research Record*, Volume 2115, pp. 127-137.
- Mulley, C., 2014. Accessibility and Residential Land Value Uplift: Identifying Spatial Variations in the Accessibility Impacts of a Bus Transitway. *Urban Studies*, 51(8), pp. 1707-1724.
- Melo, P., Graham, D., Levinson, D. & Aarabi, S., 2017. Agglomeration, accessibility and productivity: Evidence for large metropolitan areas in the US. *Urban Studies*, 54(1), pp. 179-195.
- Srour, I., Kockelman, K. & Dunn, T., 2002. Accessibility Indices: Connection to Residential Land Prices and Location Choices. *Transportation Research Record*, Volume 1805, pp. 25-34.
- Victorian Government, 2013. *Travel in Metropolitan Melbourne: Vista Survey 2013*, Melbourne: Victorian Government.
- Waddell, P., 2011. Integrated Land Use and Transportation Planning and Modelling: Addressing Challenges in Research and Practice. *Transport Reviews*, 31(2), pp. 209-229.