City proximity, travel modes and house prices: The tale of three cities in Sydney

Shanaka Herath[a]* and Ajith Jayasekare[b]

[a]*School of Economics / SMART Infrastructure Facility, Building 6 Room 204, University of Wollongong, Wollongong, NSW 2522 Australia, E shanaka@uow.edu.au, T +61 2 4239 2330 ;* [b]*SMART Infrastructure Facility, University of Wollongong, Wollongong, NSW 2522 Australia*

Abstract

House prices are partly explained by proximity to the urban centre. Generally, for simplicity, proximity to the urban centre is measured via straight-line distance (i.e. the distance 'as the crow flies'). However, distance between two points in space can be mainly conceptualised in three ways – straight-line distance, road-network distance and overland distance. Therefore, the particular distance measure that portrays 'reality' as closely as possible in a given study is context-specific. We examine implications of using different measures of distance on house price analyses using a dataset for Sydney. Spatial econometric techniques provide a mechanism to compare different distance measures in a robust manner. The disaggregated analysis of three regions in the city confirms distinct distance metrics exhibit different effects on house prices. Improving the modelling procedure taking into account the local context leads to the accurate measurement of 'city centre effects', informing policy makers on the actual extent of house price decline with an additional km of distance from the city centre. A separate section links these findings to prevalent travel modes in different parts of Sydney, and it suggests there seem to be three different cities in Sydney in terms of residents' preferred travel modes and their willingness to pay for 'proximity to the city centre'. Whilst revealing how residents' preferences for transport modes are reflected in house prices, these applications can guide in planning transport infrastructure projects (e.g. roads, highways and walking paths), establishing speed limits and in improving public transport efficiency.

Key in the subject of the topic (selecting from the list posted under "Call for Papers" on the website of prres2018.nz): *Urban economics*

City proximity, travel modes and house prices: The tale of three cities in Sydney

House prices are partly explained by proximity to the urban centre. Generally, for simplicity, proximity to the urban centre is measured via straight-line distance (i.e. the distance 'as the crow flies'). However, distance between two points in space can be mainly conceptualised in three ways – straight-line distance, road-network distance and overland distance. Therefore, the particular distance measure that portrays 'reality' as closely as possible in a given study is context-specific. We examine implications of using different measures of distance on house price analyses using a dataset for Sydney. Spatial econometric techniques provide a mechanism to compare different distance measures in a robust manner. The disaggregated analysis of three regions in the city confirms distinct distance metrics exhibit different effects on house prices. Improving the modelling procedure taking into account the local context leads to the accurate measurement of 'city centre effects', informing policy makers on the actual extent of house price decline with an additional km of distance from the city centre. A separate section links these findings to prevalent travel modes in different parts of Sydney, and it suggests there seem to be three different cities in Sydney in terms of residents' preferred travel modes and their willingness to pay for 'proximity to the city centre'. Whilst revealing how residents' preferences for transport modes are reflected in house prices, these applications can guide in planning transport infrastructure projects (e.g. roads, highways and walking paths), establishing speed limits and in improving public transport efficiency.

Keywords: Euclidean distance; road-network distance; overland distance; house price; travel mode; spatial econometrics

Introduction

Most economic activities and social exchange in a city occur in the urban city centre. Building on Thünen model (1826) of city structure and land use patterns (Thünen, 1966) and Alonso's seminal work on the monocentric model (Alonso, 1964), proximity to city centre is a key determinant of urban house prices. Subsequent contributions of Mills and Muth expanded this notion to what is currently known as a more general version of the monocentric model (Mills, 1967; Muth, 1969). The applications of this theory since the 1970s have firmly established that land values are partly determined by distance from the city centre (Karlsson, 2011). However, with the emergence of multi-centric cities, this theory has been extensively critiqued by opponents due to its 'deviation from the reality'. The scholarly response to these criticisms has been to expand the monocentric model to include urban sub-centres (Brueckner, 2011).

Regardless of whether distance(s) from a single urban centre or multiple centres are included within house price models, an important consideration is how to accurately measure distances from the city centre(s) to housing locations. Generally, for simplicity, distance 'as the crow flies' is included without much attention to the empirical context at hand. However, distance between two points in space is mainly conceptualised in three different ways – straight-line (Euclidean) distance (distance 'as the crow flies'), vector-based[1] road-network distance ('travel distance') and raster-based[2] road distance ('overland distance'). Therefore, the particular distance measure that portrays 'reality' as closely as possible in a given study is context-specific.

How do we know what distance measure portrays the 'reality' as closely as possible? If the most appropriate distance measure is not used, this may either result in this variable being statistically insignificant in the model or the model being below par in terms of explanatory power. Following examples demonstrate few such scenarios:

- If road-network data does not include footpaths and 'shortcuts'[3] that people use instead of main roads, the straight-line distance may be most representative

---

[1] A vector has magnitude (i.e. how long it is) and direction.
[2] In a raster map, a vector or an image is converted into pixels that can be used for computations (e.g. travel time calculations).
[3] Many cities promote walking by making information available to public on footpaths and shortcuts (see for example http://www.cityofsydney.nsw.gov.au/explore/getting-around/walking/sydney-walks/redfernchippendale-shortcuts).

- If most of the commuting takes place via the road-network and actual distances travelled vary substantially from straight-line distances due to topography, existing built environment or other factors unique to the study area, road-network distances are more likely to be representative
- If distances and speed limits are highly variable, a more sophisticated distance measure such as overland distance may be warranted

The first scenario is most likely within inner city areas, the second in the middle suburbs and the third in fringe areas.

The flaws of house price models problematically skew our understanding of urban structure and spatial distribution of amenities (Dziauddin et al., 2014), and misinformed policy interventions and poor infrastructure planning decisions result (Herath, 2016). This paper proposes to consider different distance metrics as part of the house price modelling procedure. Identifying the most appropriate distance metric facilitates the accurate measurement of ‘city centre effects' and informs policy makers on the actual extent of price decline of an additional km of distance away from the city centre. In addressing the concern that *omitted variable problem* may result in spurious estimates (Kuminoff et al., 2010), the analysis is extended using spatial econometric techniques that enable a comparison of different distance measures in a robust manner. Based on this framework, implications of using different measures of distance on model estimates are compared for Sydney. Spatial models also reveal spatial relationships (hereafter, spatial effects) of house prices that exist within the study area. The detailed analysis unpacks the possible link between statistically significant distance variables and prevailing travel modes in different regions in the city. This information is useful in planning infrastructure projects (e.g. roads, highways and footpaths), establishing speed limits, and in improving public transport efficiency.

The paper is structured as follows: the next section discusses different definitions of distance as variables measuring locational effects of house prices. Section 3 presents the methodology, including the standard hedonic model, diagnostics to test for presence of spatial effects and the spatial hedonic model. Data and variables are described in Section 4. The main results in Section 5 illustrate aggregate and region-specific findings of this research and also reflect on the possible link between the findings and location-specific travel modes. Section 6 closes with some policy implications.

Distance from the city centre as a variable measuring locational effects of house prices

The 'quality' of a house's location could be explained either referring to its *absolute location* or *relative location*. The former is a concept that expresses the location of a house in terms of longitude and latitude (XY) coordinates so that all locational features in close proximity (also measured in XY coordinates) can be identified. This is operationalised in research testing *urban amenities theory* that explains the impact of nearby amenities on house prices. Examples of amenities shown to influence house prices include school quality (Nguyen-Hoang and Yinger, 2011), urban forests (Zygmunt and Gluszak, 2015), public parks and open spaces (McCord et al., 2014), scenic view (Baranzini and Schaerer, 2011), air quality (Zheng et al., 2013) and crime rate (Buonanno et al., 2012). Typically, absolute location metrics measure proximities to local amenities, and a common approach to defining such variables is to assign distance bands. For example, houses located within a 400 metre band from a bus stop are coded '1' within a categorical variable so the effect of proximity to a bus stop can be evaluated.

Relative location refers to the location of a house relative to another location: in terms of house prices, this often is the location of urban centre. This definition is common in urban economics where distance from, or access to, the CBD is a primary determinant of house prices. The tradition of incorporating relative distance dates back to the Thünen model (1826) which considered distance from the city centre in a study of agricultural land use. The monocentric model of Alonso (1964) included the same variable in a study of urban regions, and the AMM model[4] (1967-1972) that revised and generalized the monocentric model are subsequent applications of the distance variable defined in that way. In its present form, monocentricity suggests house prices are influenced by distance from the CBD, transportation costs, household income, metro-area population size and agricultural rental rates. Relative distances from multiple urban centres are incorporated within multi-centric models.

Due to long distances involved, the impact of proximity to CBD (i.e. relative distance) is often measured by computing distance in kms. This is in contrast to distance bands (see above) used to capture effects of local amenities (i.e. absolute distance). There are three main approaches to calculate distances between spatial points – straight-line distance (Euclidean distance), vector-based road-network distance ('travel distance') and raster-based road distance ('overland distance') (Nilsson, 2014; Karlsson, 2011; Sander et al., 2010).

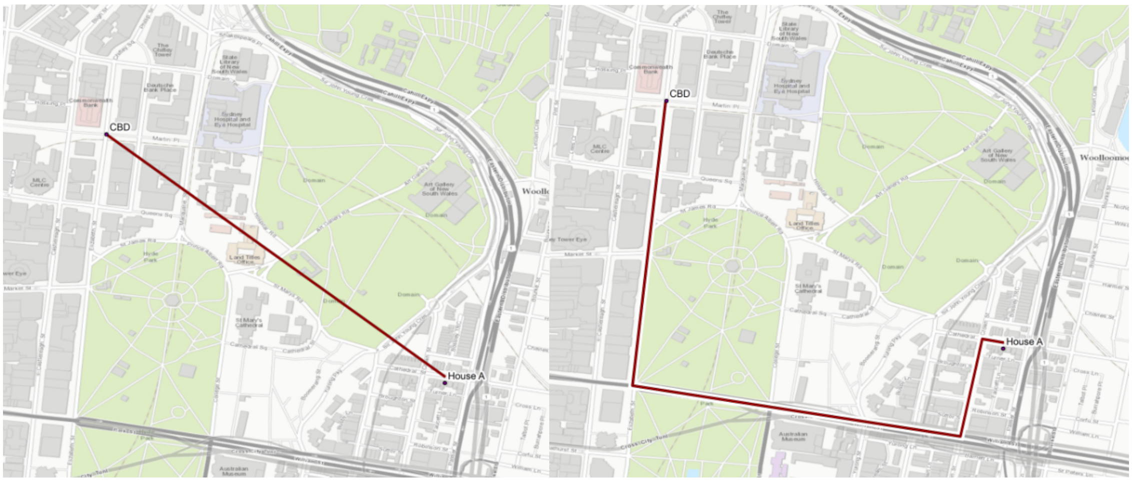
Fig. 1 (A) illustrates straight-line distance (Euclidean distance) in 2D (two-dimensional) plane, also known as the Aerial distance. In simple terms, this is the

---

[4] The Alonso-Muth-Mills model

distance ‘as the crow flies’. Distances used in house price models are generally understood as straight-line distances unless specified otherwise, mainly because they are easy to understand conceptually and simple to calculate. However, critics have questioned the usefulness of this metric as people generally travel along roads, train tracks or walking paths (Lu et al., 2011).

Fig. 1 – (A) Straight-line distance in 2D plane (B) Road-network distance (travel distance)



(A) (B)

Source: Authors own work

The road-network distance, also known as travel distance, is illustrated in Fig. 1 (B). This is computationally intensive and requires greater user effort and knowledge, and it allows addition of impedance (i.e. road types or one-way streets) and modality (i.e. walking, cycling or driving). This essentially is the distance travelled from home to an amenity by the average person using any travel mode. There may be alternative routes to reach the same destination however the shortest path among available routes is computed. Road-network distance is increasingly been used to represent access to amenities.

The third metric overland distance is calculated based on travel times, taking into account impedances or weights, or alternatively, travel costs, computed via a cost-weighted distance function. A vector road network is rasterised (see footnote 2 on page 3) and the shortest path between two points is calculated. This can also be illustrated using Fig. 1 (B). For instance, rather than measuring road distance in kms, this metric calculates the travel time to the CBD based on speed limits at particular sections of a route. An assumption here is that the travel speed is the maximum allowable at specific sections of the route. The functional form of the travel time calculation is as follows:

$$Travel\ time\ = \sum (Length\ /Speed\ limit)$$

Despite the fundamental differences between these definitions of distance, they are arbitrarily employed in previous studies. Choosing a distance metric should however be an important consideration within the modelling procedure as the particular distance measure that most appropriately represents proximity in a given study is specific to that context. If the most appropriate distance measure is not used, this may either result in this variable being statistically insignificant in the model or the model may exhibit lower explanatory power. In this paper, we present a strong case for the different distance measures to be considered as part of an assessment of how these variables behave in specific city contexts.

Methodology

The central analytical methodology used in this research is hedonic price method (HPM). HPM considers the value of a house to be equal to the sum of the implicit values of utility-bearing housing characteristics:

$$P = X\beta + \varepsilon$$

where, P is $n \times 1$ vector of house prices, X is $n \times j$ matrix of housing characteristics, β is $j \times 1$ vector of coefficients associated with housing characteristics, and ε is $n \times 1$ vector of random error terms that is IID[5] $(0, \sigma^2)$.

The housing characteristics (X) include structural attributes, neighbourhood characteristics and locational features of houses. Nearby amenities are incorporated as locational features so the potential values that home-buyers assign to specific amenities can be appraised. The main variable of interest – i.e. distance from the CBD – is included within locational features.

Improving the model via incorporation of spatial effects

The standard hedonic regression model has however faced criticism mainly due to the omission of spatial effects (LeSage and Pace, 2009). These spatial properties arise when points – i.e. houses in our context – are close to each other, as values observed at a location tend to depend on values of nearby locations. The critics, beginning with Tobler (1970) who famously published 'all places are related but nearby places are more related than distant places', have helped establish a research tradition integrating spatial effects into house price models. Herath and Maier (2013) and Wilhelmsson (2002) demonstrate *spatiality* should be a critical consideration in house price analyses as spatial autocorrelation could arise from numerous sources:

(1) Price of a house is affected by the prices of neighbouring houses;
(2) Relevant spatially correlated variables have been omitted; or
(3) Functional form is misspecified or suffers from measurement error.

[5] The traditional assumption of independent and identically distributed (i.i.d.) variable

The concerns about spatial effects are systematically addressed within this research. After estimating the hedonic model via ordinary least squares (OLS), statistical tests are undertaken to examine spatial autocorrelation within data. Spatial autocorrelation is the formal property that measures the degree to which near and distant things are related[6]. Statistical tests of match between locational similarity and attribute similarity – e.g. Global Moran's I[7] – highlight the presence of such spatial linkages. They indicate positive, negative or neutral relationships. If any extant spatial effects are not taken into account within the model, the assumption of IID errors will be violated (LeSage and Pace, 2009).

Typically, spatial linkages between houses exist within neighbourhoods. In implementing the spatial autocorrelation tests (e.g. Global Moran's I), the focus has been on the 'appropriate size' of an area that constitutes a neighbourhood. In applied work, this involves making judgements about which other houses are to be taken into account when identifying spatial linkages of a specific house. Previous research has demonstrated two main ways in which a neighbourhood can be defined based on physical distance – distance band or k-nearest neighbours. The former asks the question *what is the physical size of a neighbourhood?* This, for instance, could be a 500 metre great-circle distance from each house. The latter concerns *how many neighbours are to be included?* As an example, five closest houses of each house could be included. Once a neighbourhood criterion is determined, weights are assigned within a spatial weights

[6] The existence of such relationships within data is termed 'spatial dependence'
[7] See Anselin (1988) for more details of these tests.

matrix (SWM)[8] to represent the neighbourhood structure of houses. Spatial autocorrelation tests incorporate these spatial weights matrices.

Subsequent to a statistically significant spatial autocorrelation test, the Lagrange Multiplier (LM) tests are carried out to determine the type of spatial dependence. LM tests also include SWMs as a way to incorporate neighbourhood linkages. They check for two types of spatial dependence that occur due to dependence in spatial error λ and spatial lag ρ. The LM tests should be supplemented by their robust counterparts – an improved alternative that conducts specification testing with locally misspecified alternatives – when both LM Error and LM Lag tests are significant (see Bera and Yoon (1993) for more details). These tests guide in determining what type of spatial regression model to be estimated. Two established alternative spatial model specifications are as below:

Spatial error model incorporates spatial effects through error term

$$P = X\beta + \varepsilon$$

$$\varepsilon = \lambda W\varepsilon + \xi$$

where, $W\varepsilon$ is the spatially weighted vector of error terms, λ is the error correlation coefficient, and ξ is a vector of uncorrelated error terms. All other symbols are as previously defined. If there is no spatial correlation between the errors, then $\lambda = 0$.

Spatial lag model incorporates spatial effects by including a spatially lagged dependent variable as an additional predictor

[8] In cases where houses have an unequal number of neighbours, row standardization (W) is used to create proportional weights within spatial weights matrices. W weights increase the influence of links from observations with few neighbours. Note the possibility of houses with no neighbours.

$$P = \rho W p + X\beta + \varepsilon$$

where, Wp represents spatially lagged dependent variables and ρ is the spatial autoregression coefficient. All other symbols are as previously defined. If there is no spatial dependence, and P does not depend on neighbouring p values, $\rho = 0$.

Addressing the omitted variable bias

The ‘missing location variables’ is a well-documented problem in house price modelling (Kuminoff et al., 2010; Abbott and Klaiber, 2010). Though we have included a long list of locational variables within the models (see Section 4), the following additional measures serve to further eliminate any remaining influence of possible ‘omitted variable bias’:

The **first strategy** is to focus on a narrow geographic area where many influences are already controlled for (e.g. see Brasington 2004). We focus on Sydney metropolitan area whilst excluding diverse regional and rural housing markets outside of Sydney and within New South Wales (NSW).

However, intra-regional differences within Sydney are also considerable. Therefore, following previous research (e.g. see Beron et al 2001, Deaton 1988), the **second strategy** is to utilise fixed-effects variables – i.e. categorical variables for unique Local Government Areas (LGAs) – to capture the influence of omitted variables within a spatial context. The locational effects associated with missing local particularities including neighbourhood characteristics, amenities such as environmental and recreation facilities, access to public transport and service centres, social composition and other omitted spatial variables such as local demand determinants are captured by local fixed-effects (dummy) variables (Herath and Maier, 2013). Our fixed-effects variables also control for missing structural variables, as structural characteristics are likely to be similar within nearby local neighbourhoods (Orford, 1999).

As a **third strategy**, above mentioned spatial models also address the omitted variable problem – "Spatial statistics represent a powerful, underutilized tool in urban and environmental economics, capable of addressing omitted variable bias" (Brasington & Hite 2005, p. 58).

When using a *spatial lag* model, similar to the way a time lag of the dependent variable picks up unobserved autoregressive influences, the spatial lag term picks up unobserved influences and omitted variables, both property-specific as well as related to neighbouring properties that affect house value (Bolduc et al., 1995; Griffith, 1988):

> Unmeasured influences help determine the value of neighboring houses and … the value of neighboring houses is related to the value of our own house. So our own house value is affected by the unmeasured influences of neighboring observations. And the unmeasured influences of neighboring houses are similar to the unmeasured influences for our house because our neighbors are close: the same things that affect our neighbors should affect us, too. So the spatial lag dependent variable incorporates the influence of omitted variables on the value of our own house (Brasington & Hite 2005, p. 63).

In contrast, the *spatial error* model implies a situation where omitted variables follow a spatial structure such that the error variance-covariance matrix is no longer diagonal. Spatial fixed-effects (see above) may address this problem at least partly, nevertheless empirical studies have shown this does not resolve the issue completely so the need for the spatial error model (Anselin and Lozano-Gracia, 2007). Additionally, spatial models also allow capturing omitted effects at a more localized level than by including fixed-effects variables.

The originality of the present study is that it employs spatial analysis techniques that are capable of capturing more influence of omitted variables than the standard models. The spatial effects may include omitted variables or other forms of spatial dependence that are present in housing markets. Whatever the cause of spatial effects, the paper presents a strong case for the spatial models to be used to increase accuracy and to reduce the bias of parameter estimates.

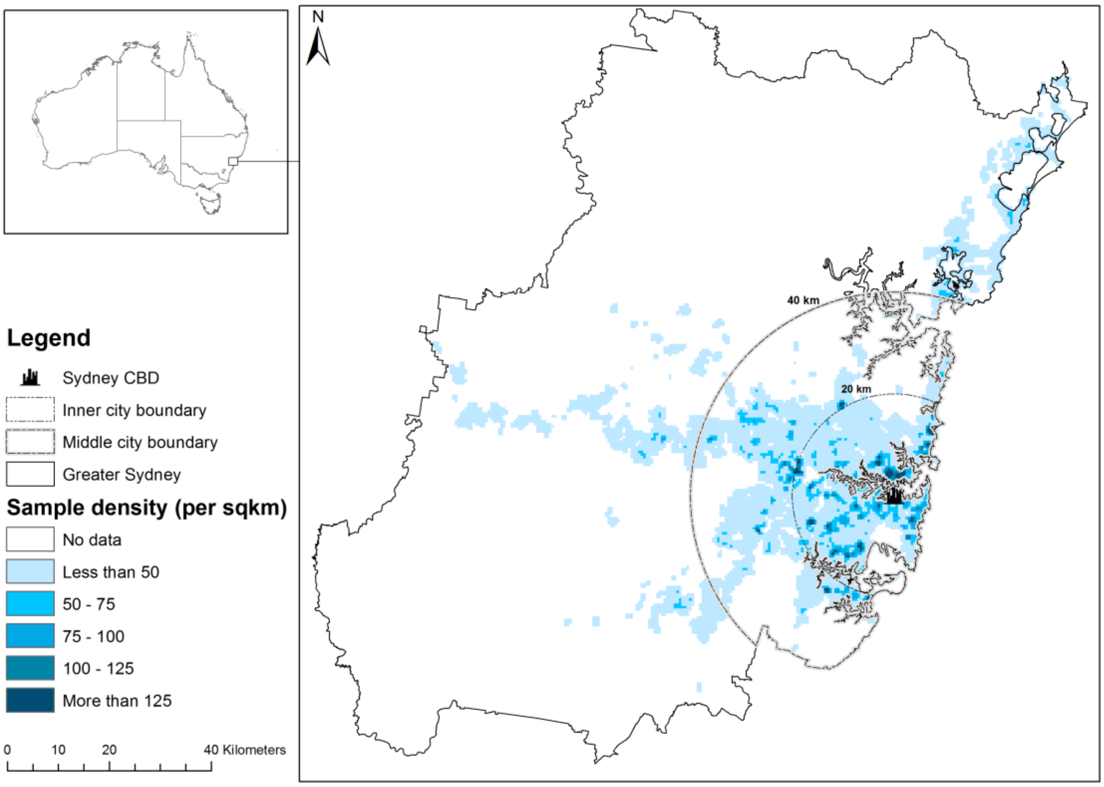
Data and variables

The dataset is assembled from multiple sources. House prices, housing locations, transaction details and structural attributes are obtained from Australian Property Monitors (APM) via the Australian Urban Research Infrastructure Network (AURIN). This covers all houses[9] sold in Greater Sydney area from January to December 2011 (see Fig. 2). The Australian Bureau of Statistics (ABS) Census 2011 provides neighbourhood characteristics at local level (SA1)[10]. A list of top 100 schools in NSW in 2010 is obtained from the website http://bettereducation.com.au/, which ranks private, public and selective high schools based on the percentage of distinguished achievers (D.A.) in HSC (High School Certificate) examinations[11]. The addresses of schools are geocoded to be able to map the best schools across our study area. Locations of bus stops and train stations are sourced from the NSW open data portal

[9] Only houses are considered in this study to retain a homogeneous sample. This strategy also simplifies the spatial regression analysis due to absence of multiple dwellings with the same XY coordinates.
[10] Statistical Area Level 1 (SA1) is the smallest geographical unit for the release of census data with an average population size of approximately 400 people.
[11] This is achieved as follows: HSC results are divided statistically into bands, from 1 (lowest) to 6 (highest) in a standard course, or from E1 to E4 in an extension course. HSC Distinguished Achievers lists the students who achieved a result in the highest band (Band 6 or Band E4) for one or more courses. This is equivalent to an HSC score of 90% or more. It is Australian Tertiary Admission Rank (ATAR) score related. The school reported is the student's main school – in some cases the student may have studied one or more courses at another school.

http://data.nsw.gov.au/. Geographic Information System (GIS) data on locations of parks, beaches and main roads are generated using the Open Street Map (OSM).

Fig. 2 Locations of sample houses in Sydney



Source: Based on APM housing transactions, ABS digital boundaries in 2011

For locational amenities typically farther from houses (i.e. parks and beaches), road-network distance was used. Proximities to bus stops, train stations and major roads were calculated based on straight-line distances to be able to capture the associated environmental externalities such as pollution and noise. This judgment was also influenced by the fact that walking to nearby bus stops / train stations often involves using footpaths or 'shortcuts' however our road network didn't include footpaths. The categorical variable representing access to bus stops includes three distance bands. The first captures houses located within 100 metres of bus stops to investigate the assertion of negative externalities arising from noise and pollution – e.g. Seo et al. (2014). The second includes houses located outside this 100 metre boundary but within a 400 metre band from a bus stop, to assess the positive externalities associated with proximity to public transport. The remaining band (>400m) should demonstrate a smaller or no price premium compared to the 100-400m band. Given the potentially large area-effects of train stations compared to bus stops, the corresponding distance bands for train stations were increased to 'less than 200 metres', 'between 200 – 800 metres' and 'more than 800 metres'.

Table 1 Descriptive statistics

Variable	Description	Mean	Median	Min	Max
price	Price (AUD)	725,754	560,000	113,000	18,500,000
evnt_ty	Event type (auction = 9.5%, private sale = 90.5%)				
Structural variables					
prop_typ	Property type (house = 85.2%, cottage= 0.3%, semi-detached = 1.0%, terrace = 0.2%, townhouse = 13%, villa = 0.4%)				
bedr	Number of bedrooms	3.4	3	1	10
baths	Number of bathrooms	1.8	2	1	9
park	Number of parking spaces	1.8	2	1	10
Locational variables					
lga	Local Government Area (Ashfield = 0.5%, Bankstown = 4.4%, Blacktown = 8.7%, Blue Mountains = 2.7%, Burwood = 0.4%, Camden = 2.3%, Campbelltown = 4.2%, Canada Bay = 1.3%, Canterbury = 2.2%, City of Auburn = 1.2%, City of Kogarah = 1.1%, Fairfield = 3.1%, Gosford = 6.0%, Hawkesbury = 1.1%, Holroyd = 2.3%, Hornsby = 3.6%, Hunters Hill = 0.3%, Hurstville = 1.7%, Ku-Ring-Gai = 3.0%, Lane Cove = 0.6%, Leichhardt = 1.0%, Liverpool = 4.4%, Manly = 0.7%, Marrickville = 0.9%, Mosman = 0.5%, North Sydney = 0.7%, Parramatta = 3.6%, Penrith = 5.1%, Pittwater = 1.7%, Randwick = 1.6%, Rockdale = 1.7%, Ryde = 2.4%, Strathfield = 0.6%, Sutherland Shire = 5.5%, Sydney = 0.7%, The Hills Shire = 4.9%, Warringah = 2.9%, Waverley = 0.7%, Willoughby = 1.4%, Wollondilly = 0.2%, Woollahra = 0.7%, Wyong = 5.5%)				
d2k_sch	1 if a top school is located within 2 kms (road-network distance), 0 otherwise		0	0	1
n_pk	Road-network distance to the closest public park (in meters)	1,317.2	949.5	0.3	37,257.1
n_bch	Road-network distance to the closest beach (in meters)	21,238.4	19,351.5	0.3	102,824.7
c_e_bus	Straight-line distance from the closest bus stop (<100m = 8.9%, 100-400m = 86.4%, >400m = 4.8%)				

c_e_train	Straight-line distance from the closest train station (<200m = 0.5%, 200-800m = 15.2%, >800m = 84.2%)				
e_mjrd	Straight-line distance from the closest major road (in meters)	947.6	546.0	0.7	18,446.5
e_syd	Straight-line distance from Sydney (in meters)	28,751.5	24,340.7	903.1	95,018.5
n_syd	Road-network distance from Sydney (in meters)	35,993.1	29,037.3	1,099.0	118,898.7
o_syd	Overland distance from Sydney (in seconds)	1,431.0	1,171.9	68.3	5,109.3
Neighbourhood variables (calculated for SA1s)					
une	per cent of unemployed persons	5.7	5.1	0.0	50.0
inc	per cent of people with weekly income > $1500	16.3	14.3	0.0	60.0
age_prop	per cent of people aged > 60	18.4	17.8	0.0	96.4
net_mig	per cent of people born outside of Australia	31.7	30.4	0.0	100.0
density	persons per square km	3,259.9	3,004.4	2.2	30,059.3

Source: Author calculations
Notes: Records with the top and bottom 1% of prices and structural variables were excluded to neutralise the effects of outliers. Percentages are shown for each categorical variable (see description); sample size = 39199 observations.

The descriptive statistics are presented in Table 1. About 10 per cent of houses in the sample are sold in auctions and the average price is approximately $726,000. The median price is lower at $560,000 due to high variability of prices at the top end of the market – i.e. the presence of multi-million-dollar houses. The sample comprises a majority of detached houses (85 per cent) and townhouses (13per cent). The ‘average house’ has three bedrooms, two bathrooms and two parking spaces.

Our primary research question requires computing distances from the city centre to houses based on three different definitions of distance – straight-line distance, road-network distance and overland distance. The calculated straight-line distances between the CBD and houses range from 900 meters to 95 kms and road-network distances from 1.1 km to 119 kms. Overland distances, expressed in travel time, range from 68 seconds to 1 hour and 40 minutes. All 42 Local Government Areas (LGAs) in Greater Sydney are represented within the sample. On average, a park is located 1.3 kms, a beach 21 kms and a major road 950 meters away from houses. The small-scale neighbourhoods in the sample comprise on average 6 per cent of unemployed persons, 16 per cent with high incomes (defined here as earning more than $1,500 p/w), 18 per cent aged above

60 years and 32 per cent born outside of Australia. Neighbourhood density revealed a vast variability between 2 to 30,000 persons per square km, with an average of 3,300 persons per square km.

Results

In preparation for the estimation of the hedonic model, the bivariate correlations within the dataset were calculated. The initial expectation was to include distances from both Sydney and Parramatta in the model because the latter is increasingly considered as an important urban centre in NSW: since 2000, Parramatta has emerged as a government centre, a major business and commercial hub, and the second largest CBD in Sydney. However, the correlation analysis indicated 'distance from Sydney CBD' is strongly correlated with 'distance from Parramatta CBD' and thus the latter was excluded from the analysis. This strong positive correlation is likely to have resulted from the close proximity between Parramatta and Sydney within the context of Sydney Metro area.

As expected, straight-line, road-network and overland distances from Sydney were also correlated. Given our focus on effects of different distance variables – i.e. straight-line, road-network and overland distances – on house prices, and also due to the fact that these variables are correlated, three versions of the model were estimated. Each version included distance from the CBD measured based on one of the above definitions, alongside other variables. This strategy addresses the potential problem of multicollinearity. The results of the standard hedonic model are presented in Table 2:

Table 2 Estimation results for the variants of the standard model

Variable	Standard *Model 1*[a]	Standard *Model 2*[b]	Standard *Model 3*[c]
Constant	15.339***	15.232***	14.434***
	(0.076)	(0.082)	(0.064)

d2k_sch	0.012***	0.011**	0.012***
	(0.003)	(0.003)	(0.003)
lga.Ashfield	0.102***	0.080***	0.021
	(0.020)	(0.020)	(0.020)
lga.Bankstown	-0.127***	-0.189***	-0.270***
	(0.019)	(0.019)	(0.018)
lga.Blacktown	-0.208***	-0.270***	-0.371***
	(0.021)	(0.022)	(0.021)
lga.Blue Mountains	-0.087***	-0.164***	-0.287***
	(0.026)	(0.027)	(0.026)
lga.Burwood	0.168***	0.133***	0.067**
	(0.022)	(0.022)	(0.022)
lga.Camden	-0.173***	-0.249***	-0.357***
	(0.024)	(0.025)	(0.023)
lga.Campbelltown	-0.315***	-0.382***	-0.483***
	(0.023)	(0.023)	(0.022)
lga.Canada Bay	0.214***	0.205***	0.147***
	(0.017)	(0.017)	(0.017)
lga.Canterbury	-0.007	-0.055**	-0.121***
	(0.017)	(0.018)	(0.017)
lga.City of Auburn	-0.049*	-0.095***	-0.172***
	(0.020)	(0.020)	(0.019)
lga.City of Kogarah	0.101***	0.028	-0.047*
	(0.020)	(0.020)	(0.019)
lga.Fairfield	-0.120***	-0.188***	-0.277***
	(0.021)	(0.022)	(0.021)
lga.Gosford	-0.281***	-0.299***	-0.425***
	(0.025)	(0.027)	(0.026)
lga.Hawkesbury	-0.077**	-0.143***	-0.249***
	(0.025)	(0.026)	(0.025)

lga.Holroyd	-0.159***	-0.223***	-0.315***
	(0.020)	(0.021)	(0.020)
lga.Hornsby	0.025	-0.019	-0.102***
	(0.019)	(0.020)	(0.019)
lga.Hunters Hill	0.367***	0.376***	0.317***
	(0.022)	(0.023)	(0.022)
lga.Hurstville	0.020	-0.039*	-0.109***
	(0.019)	(0.019)	(0.019)
lga.Ku-Ring-Gai	0.273***	0.232***	0.167***
	(0.018)	(0.018)	(0.017)
lga.Lane Cove	0.260***	0.274***	0.240***
	(0.018)	(0.019)	(0.019)
lga.Leichhardt	0.079***	0.107***	0.079***
	(0.016)	(0.016)	(0.016)
lga.Liverpool	-0.216***	-0.293***	-0.392***
	(0.022)	(0.022)	(0.021)
lga.Manly	0.433***	0.443***	0.375***
	(0.019)	(0.019)	(0.019)
lga.Marrickville	-0.044**	-0.067***	-0.110***
	(0.017)	(0.017)	(0.017)
lga.Mosman	0.647***	0.679***	0.634***
	(0.019)	(0.019)	(0.019)
lga.North Sydney	0.222***	0.236***	0.211***
	(0.017)	(0.017)	(0.017)
lga.Parramatta	-0.088***	-0.135***	-0.222***
	(0.019)	(0.020)	(0.018)
lga.Penrith	-0.182***	-0.264***	-0.376***
	(0.023)	(0.024)	(0.022)
lga.Pittwater	0.434***	0.396***	0.299***
	(0.021)	(0.022)	(0.020)

lga.Randwick	0.306***	0.264***	0.227***
	(0.016)	(0.016)	(0.016)
lga.Rockdale	-0.031.	-0.092***	-0.160***
	(0.018)	(0.018)	(0.018)
lga.Ryde	0.129***	0.102***	0.033*
	(0.017)	(0.018)	(0.017)
lga.Strathfield	0.230***	0.189***	0.117***
	(0.021)	(0.021)	(0.020)
lga.Sutherland Shire	0.009	-0.056**	-0.145***
	(0.020)	(0.020)	(0.019)
lga.The Hills Shire	-0.036.	-0.084***	-0.175***
	(0.020)	(0.021)	(0.020)
lga.Warringah	0.202***	0.184***	0.111***
	(0.017)	(0.018)	(0.017)
lga.Waverley	0.553***	0.543***	0.513***
	(0.017)	(0.018)	(0.018)
lga.Willoughby	0.277***	0.255***	0.215***
	(0.016)	(0.016)	(0.016)
lga.Wollondilly	-0.145***	-0.227***	-0.329***
	(0.031)	(0.031)	(0.031)
lga.Woollahra	0.706***	0.718***	0.704***
	(0.017)	(0.017)	(0.017)
lga.Wyong	-0.318***	-0.378***	-0.506***
	(0.027)	(0.029)	(0.027)
prop_typ.cottage	0.051**	0.050*	0.049*
	(0.019)	(0.019)	(0.019)
prop_typ.semi	-0.061**	-0.064**	-0.065**
	(0.022)	(0.022)	(0.022)
prop_typ.terrace	-0.074*	-0.072*	-0.066*
	(0.031)	(0.031)	(0.031)

prop_typ.townhouse	-0.180***	-0.183***	-0.184***
	(0.019)	(0.020)	(0.020)
prop_typ.villa	-0.090***	-0.094***	-0.095***
	(0.025)	(0.025)	(0.025)
evnt_ty	0.033***	0.032***	0.032***
	(0.004)	(0.004)	(0.004)
LOG_bedr	0.291***	0.290***	0.290***
	(0.006)	(0.006)	(0.006)
LOG_baths	0.217***	0.218***	0.218***
	(0.003)	(0.003)	(0.003)
LOG_park	0.081***	0.081***	0.081***
	(0.003)	(0.003)	(0.003)
LOG_une	-0.005*	-0.004.	-0.004.
	(0.002)	(0.002)	(0.002)
LOG_inc	0.177***	0.180***	0.181***
	(0.003)	(0.003)	(0.003)
LOG_age_prop	0.067***	0.069***	0.070***
	(0.003)	(0.003)	(0.003)
LOG_net_mig	0.035***	0.036***	0.038***
	(0.004)	(0.004)	(0.004)
LOG_density	-0.024***	-0.023***	-0.023***
	(0.002)	(0.002)	(0.002)
LOG_e_syd	-0.248***		
	(0.007)		
LOG_n_syd		-0.222***	
		(0.008)	
LOG_o_syd			-0.201***
			(0.008)
c_e_bus_d1	0.025***	0.025***	0.025***
	(0.004)	(0.004)	(0.004)

c_e_bus_d2	0.042***	0.042***	0.043***
	(0.006)	(0.006)	(0.006)
c_e_train_d1	0.017	0.018	0.019
	(0.014)	(0.015)	(0.015)
c_e_train_d2	-0.007	-0.005	-0.005
	(0.014)	(0.014)	(0.014)
LOG_n_pk	0.007***	0.009***	0.009***
	(0.001)	(0.001)	(0.001)
LOG_n_bch	-0.078***	-0.088***	-0.090***
	(0.002)	(0.002)	(0.002)
LOG_e_mjrd	0.017***	0.020***	0.024***
	(0.001)	(0.001)	(0.001)
Multiple R-squared	0.87	0.86	0.86
Adjusted R-squared	0.87	0.86	0.86
F-statistic	3949	3912	3893
Prob (F-stat)	0.000	0.000	0.000
No. of observations	39199	39199	39199

Notes: The dependent variable is the natural logarithm of price. The reference (omitted) categories not listed for the categorical variables include 'Sydney' for LGA, 'house' for property type, 'less than 100m' for the categorical variable 'straight-line distance from the closest bus stop' (c_e_bus), 'less than 200m' for the categorical variable 'straight-line distance from the closest train stop' (c_e_train). Binary dummy variable 1 indicates presence of a certain amenity for the variables d2k_sch (a top school is located within 2 kms) and evnt_ty (sold via an auction). *** Significant at 0.1%, ** significant at 1% and * significant at 5%. Standard errors are in parentheses.
[a] Standard model specification with straight-line distance measuring CBD effects
[b] Standard model specification with road-network distance measuring CBD effects
[c] Standard model specification with overland distance measuring CBD effects
Source: Authors' calculations

Column 2 reports the results of the model that includes straight-line distance from the CBD and other control variables. We observe a negative price gradient, and the constant-quality house price is estimated to decline by 0.25 per cent with a per cent increase in distance from the CBD. Column 3 presents the estimated effect of road-network distance from the CBD on house prices. A per cent increase in road-network distance from the CBD is estimated to be associated with a 0.22 per cent decrease in house price. As presented in column 4, an increase of 1 per cent in overland distance is associated with a decline of 0.20 per cent in house price. These results are highly statistically significant at the 0.1 per cent level, and indicate all three distance metrics are strongly related to house prices.

In terms of control variables, all three models report significant positive effects of structural attributes number of bedrooms, bathrooms and parking places. Compared to detached houses, semi-detached houses, terraces, townhouses and villas are sold at a discount but not cottages. Houses sold at an auction are more expensive relative to those sold via a private treaty. A higher proportion of unemployed persons in the neighbourhood depresses house prices whereas neighbourhoods with high proportions of aged, migrant or high-income persons are associated with higher house prices. High density neighbourhoods exhibit lower prices.

Our results suggest a negative impact of locations too close to bus stops as the dummy variable representing 100-400m has a positive and significant coefficient, measured compared to the <100m distance band (i.e. the omitted category). The larger premium for houses located more than 400m away from bus stops suggests residents may be willing to walk longer than 400 metres to reach a bus stop (see Herath (2015)). The categorical variable representing proximity to train stations is insignificant. This is likely a result of two factors: the dominance of the automobile as a mode choice and the prevalence of driving, rather than walking, to train stations.

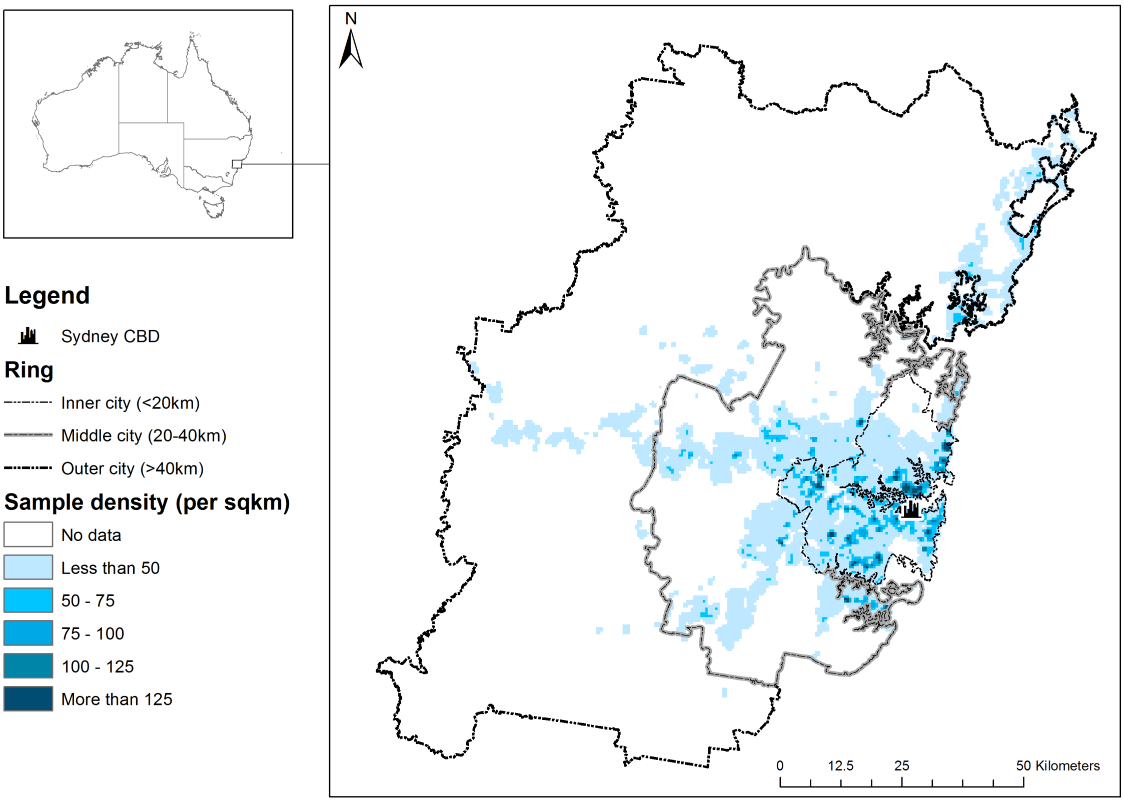
The estimates of all three models confirm the distance from a beach is a significant predictor of house prices – the farther from a beach means a lower constant-quality house price. As expected, houses located within 2kms from a highly-ranked school also exhibited a price premium. Interestingly, proximity to a major road is an off-putting factor, as the negative externalities associated with major roads (i.e. noise, air pollution and congestion) overweigh the access-related benefits. The unexpected positive effect of distance from a park on house prices is attributable to the heterogeneous nature of parks in the metro area – whilst some parks are quite attractive some others lack a proper upkeep.

Robustness check and the detailed analysis

The evidence reported above on the relationship between house prices and distance from the Sydney CBD is suggestive, but inconclusive. This is because there are different housing sub-markets in relation to distance from the CBD and numerous transport modes that are at play at different locations within Sydney, and they influence the most significant distance variables. To further investigate the assertion that the estimated effects of distance metrics on house prices are genuine, we put forward evidence on a related question: Do regions in which the houses are located within Sydney play a role in determining the most relevant distance metric? As an example, due to congestion, road conditions and/or maximum allowable speed limits associated with traveling from some areas to the CBD, it might be the case that travellers would be more concerned about travel times rather than travel distance. As such, a comparison of aggregate and region-specific findings could offer a critical interpretation of these differences and also function as a robustness test of our initial findings.

This section reports on findings of three hedonic models estimated for concentric regions within Sydney based on distance from the CBD: inner city (<20km), middle city (20-40km) and the outer city (>40km) (see Fig. 3). Similar to the aggregate model, we incorporate a range of structural, locational and neighbourhood variables as control variables. The primary results of the standard hedonic model estimated for the three regions are presented in Table 3. Panels A, B and C report the results for inner Sydney, middle Sydney and outer Sydney respectively. In each panel, rows 1, 2 and 3 present the results of the standard model estimated using straight-line distance, road-network distance and overland distance from the Sydney CBD to houses in each sub-sample.

Fig. 3 Three concentric regions in Sydney



Source: Based on APM housing transactions, ABS digital boundaries in 2011

As illustrated in Section 3, the methodology involves employing fixed-effects variables and spatial hedonic models to address omitted variable bias. Our use of categorical variables to represent unique LGAs as locational variables constitutes the former. The latter, the spatial hedonic models, incorporate the potential spatial linkages within our dataset into the model specification. These measures capture effects of missing variables.