# The value of land, floorspace, and amenities: A hedonic price analysis of property sales in Auckland 2011-2014

April 2015 Technical Report 2015/012

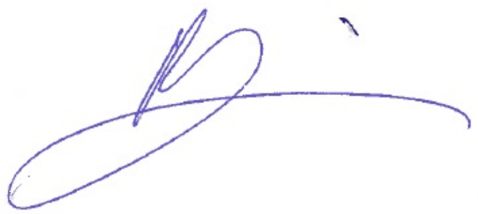
Auckland Council
Technical report 2015/012
ISSN 2230-4525 (Print)
ISSN 2230-4533 (Online)

ISBN 978-0-908320-07-3 (Print)
ISBN 978-0-908320-08-0 (PDF)

This report has been peer reviewed by the Peer Review Panel.

Submitted for review on 11 March 2015
Review completed on 21 April 2015
Reviewed by two reviewers

Approved for Auckland Council publication by:



Name: Regan Solomon

Position: Manager, Research and Evaluation Unit (RIMU)

Date: 21 April 2015

Recommended citation:
Nunns, Peter, Hitchins, Hadyn and Balderston, Kyle (2015). The value of land, floorspace and amenities: a hedonic price analysis of property sales in Auckland 2011-2014. Auckland Council technical report, TR2015/012

# The value of land, floorspace, and amenities: A hedonic price analysis of property sales in Auckland 2011-2014

Peter Nunns
Auckland Council / MRCagney

Hadyn Hitchins
Kyle Balderston
Research and Evaluation Unit
Auckland Council

With thanks to Eilya Torshizian, University of Auckland / Auckland Council and Stu Donovan, MRCagney for valuable technical advice.

# Executive summary

What makes a house – or a city – an attractive place to live? Why do people pay higher prices to live in some places than in others?

These are not idle questions. Developers ask them when choosing where and how to construct new dwellings. Individuals and families ask them when choosing which city to live in, and where to live within cities. Urban planners ask them when attempting to establish rules that govern how and where new dwellings can be developed. And, of course, economists ask them when attempting to explain the decisions that people make about housing.

This report investigates the relationship between the observed characteristics of dwellings and neighbourhoods and residential property prices. It applies a spatial hedonic price model to a dataset of recent residential property sales in Auckland. In particular, our analysis considers the following issues:

- The relative value[1] of land and floorspace to home-buyers
- The impact of location on property values – in particular, proximity to amenities such as the city centre and coastal areas
- The value that people place on other dwelling characteristics, such as pre-1940 ("heritage") status, carparking, and views of land and water
- The value that people place upon neighbourhood characteristics such as the presence of pre-1940 buildings.

Our analysis confirms some general relationships between housing prices and property features and amenities. First, we find that buyers exhibit a strong preference for more floorspace. Based on the results from our preferred hedonic price model, a spatial error regression model, we find that more living space is associated with higher sales price, as is more land. As a result, we would expect policies that enabled an increase in residential floorspace, either by enabling higher-density development or an increase in land supply for new subdivisions, to improve amenity for Aucklanders.

Secondly, we find that sales prices are strongly influenced by location within the city. People are not indifferent between different locations – all else being equal; they show a distinct preference to be closer to the city centre and a weaker, but still significant, preference to be close to the coast. It also found that after controlling for neighbourhood and location characteristics, people seemed to place higher value on older (pre-1940) buildings, but not on more car parking.

Third, our analysis also finds evidence of some "localised externalities" associated with the presence of older buildings in neighbourhoods and the preservation of coastal views. This suggests that policies that preserve these neighbourhood features may also preserve amenity for residents.

[1] Or, more precisely, the sign and statistical significance of the regression coefficients on these explanatory variables.

Fourth, we find evidence of spatial dependence in Auckland's housing market. In other words, the sale price of a single house is correlated with neighbouring property values. The Ordinary Least Squares (OLS) regression models that we tested could not fully explain these localised correlations, possibly due to omitted variables that we were not able to observe. We tested several spatial regression models, finding that a spatial error model (which treats spatial dependence as a "nuisance" to control) performed better than a spatial lag model (which treats spatial dependence as a process of interest to explain). This suggests that there are some unexplained spatial processes that influence property prices. However, this issue did not affect the overall explanatory power of our model.

Our findings also suggest that there are further opportunities to research the Auckland property market using spatial hedonic price analysis. This could include investigating the impact of infrastructure, other neighbourhood features, zoning, and special purpose overlays such as volcanic view shafts on residential property prices. In addition, applying a similar methodology to Auckland's commercial and industrial property markets could offer insights into the location preferences of Auckland firms.

# Table of contents

# 1.0 Introduction and context

What makes a house – or a city – an attractive place to live? Why do people pay higher prices to live in some places than in others?

These are not idle questions. Developers ask them when choosing where and how to construct new dwellings. Individuals and families ask them when choosing which city to live in, and where to live within cities. Urban planners ask them when attempting to establish rules that govern how and where new dwellings can be developed. And, of course, economists ask them when attempting to explain the decisions that people make about housing.

## 1.1 Theoretical background: the Rosen-Roback model

How should we interpret data on house prices?

Much of the recent literature on property prices in New Zealand has emphasised how planning regulations can drive up prices by constraining the supply of new dwellings in areas where people want to live. This literature has emphasised how binding restrictions on new dwellings can act as a "regulatory tax" that pushes up house prices above the marginal cost of production.

In an influential paper on the costs of planning regulations, Glaeser, Gyourko and Saks (2005) argue that:

> "One of the strongest implications of free markets is that in an open, competitive, unregulated market, the price of a commodity will not be greater than the marginal cost of producing that good... Free competition among these suppliers should ensure that prices are pushed down to marginal cost, so the presence of a large gap between market values and marginal production costs indicates the presence of supply-side restrictions. If we are confident that we are not missing any technological barriers to construction, then the gap between market value and the cost of supply must reflect the impact of government regulation."

A number of New Zealand-specific papers, including Grimes and Liang (2007), MRCagney (2014), Lees (2014, 2015) and Grimes and Mitchell (2015), have taken a similar perspective, modelling or estimating the degree to which specific regulations may push up dwelling prices.

However, economic theory and evidence suggests that other factors, such as the level of unpriced consumer amenity available in cities, can also result in higher property prices. The spatial equilibrium model developed by Rosen (1979) and Roback (1982) describes the trade-offs that people make between (nominal) wage levels, the cost of housing, and the level of consumer amenity available in different cities. It makes a simple yet powerful observation: that if people are willing to accept relatively low wages or high housing costs in order to live in a city, it <u>must</u> offer them benefits in order to offset these costs. The Rosen-Roback model is summarised in Table 1.

Table 1: The Rosen-Roback spatial equilibrium framework (Source: Glaeser, 2008)

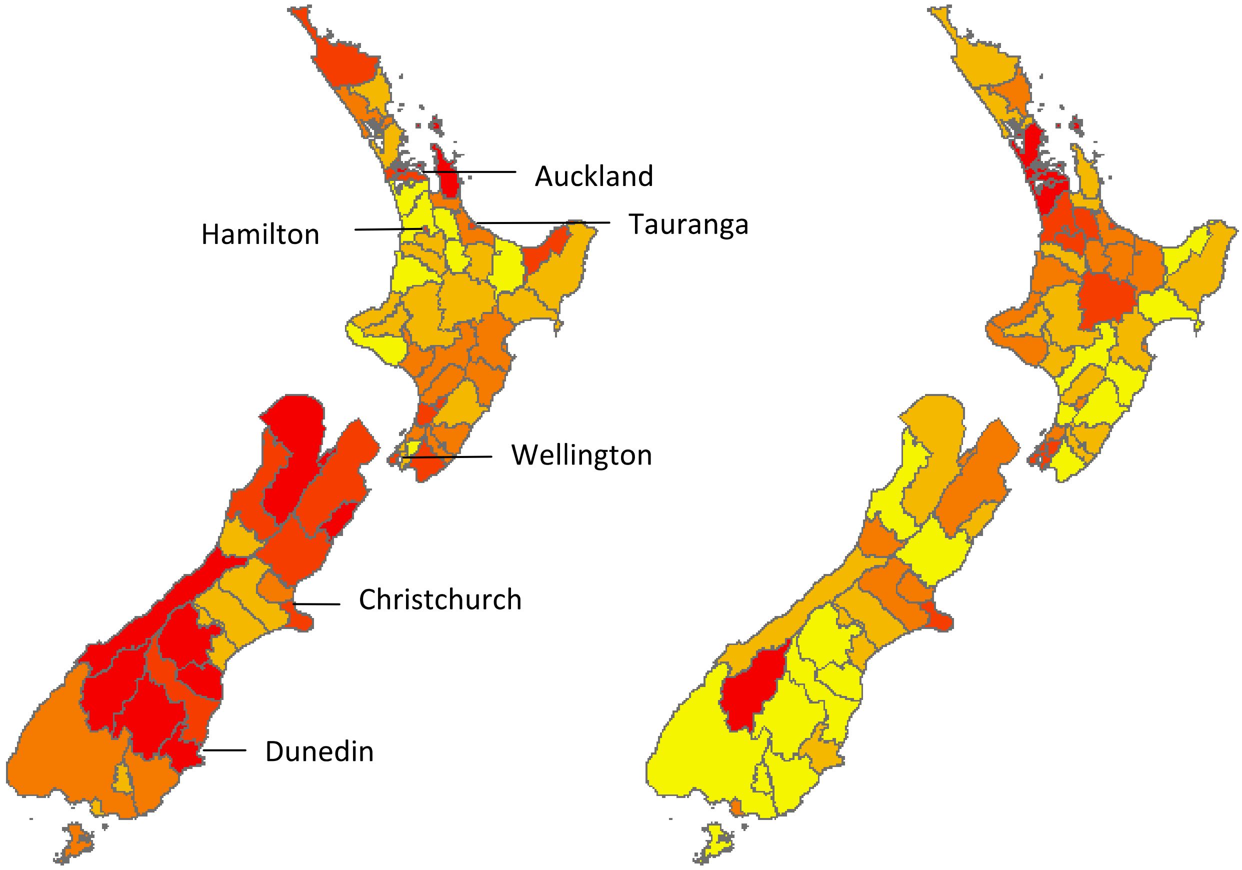
| | |
|---|---|
| Actors: | Consumers of housing |
| They choose: | Consumers choose a city that maximises their utility |
| Utility function to maximise: | U(Wages – Housing Costs, Unpriced Amenities) |
| Equilibrium condition: | U(•) is constant between locations – in equilibrium, nobody can make themselves better off by relocating |
| First order condition of utility maximisation: | $\frac{\partial(Wages - Housing\ Costs)}{\partial(Unpriced\ Amenities)} = -\frac{U_{Amenity}}{U_{Cash}}$ |

Glaeser (2008) discusses the implications of this model, noting that it indicates that "the share of income being spent on housing… is not particularly meaningful or helpful". After applying the Rosen-Roback framework to analyse population growth in US cities from 1970-2000, he observes that:

> The perverse logic of the spatial equilibrium means that <u>declining relative real wages are interpreted not as declines in productivity or well-being, but rather as rises in consumer amenities</u>. This interpretation is buttressed by the fact that these cities have had very large increases in nominal wages, uncorrected for local prices of living, but that housing costs have gone up even more. (p 65)

Several recent papers have applied this framework in New Zealand, finding that there are amenity values associated with natural and built environments and with agglomeration economies (Donovan, 2011; Grimes et al, 2014). Donovan (2011) applies the Rosen-Roback model to Census data on personal incomes and housing costs (rents) to develop an index of residential amenity in New Zealand's territorial local authorities. He interprets high housing costs within a city (relative to the city's wage levels) as an indication that it offers a higher level of amenity for residents. Figure 1 summarises his estimates of amenity, with red colours indicating higher amenity. From 1996 to 2006, the seven territorial local authories that were agglomerated into Auckland Council generally improved on the quality of life index.

Figure 1: Quality of life and quality of the business environment in New Zealand (Source: Donovan, 2011)




One implication of the Rosen-Roback model is that planning regulations that improve or preserve unpriced consumer amenities in urban areas may also result in higher housing costs relative to incomes. However, this should be interpreted as an improvement in amenity and consumer well-being, rather than a cost to society. With this in mind, planning regulations may improve amenity by:

- Managing “highly localised externalities” that may have a negative impact on residents (Anas, Arnott, and Small, 1998) and integrating uses that generate positive externalities (Chung, 1994)
- Acting as an “insurance policy against the invasion of commercial or industrial activity that would create strongly negative effects” (McDonald and McMillen, 2003)
- Providing public goods such as open space and “attractive” external built form (Chung, 1994).

## 1.2 Aim of this report

This report draws upon the Rosen-Roback framework in its analysis of Auckland house prices. It interprets housing prices as a potential indicator of amenity for residents, and asks: <u>What characteristics of dwellings and neighbourhoods are associated with higher property prices?</u> In doing so, it does not explicitly investigate the degree to which Auckland’s house prices reflect regulatory constraints on new dwelling supply or, alternatively, amenities generated by planning regulations.

This analysis can help us to understand what factors people value when purchasing properties – or, in economic jargon, the hedonic structure of property prices. In particular, we investigate several factors of interest:

- The relative value[2] of land and floorspace – previous analysis (e.g. New Zealand Productivity Commission, 2012) has considered the cost of land as an input to production of housing but has implicitly assumed that there is a fixed relationship between land and floorspace
- The impact of location on property values – in particular, proximity to amenities such as the city centre and coasts
- The value that people place on other dwelling characteristics, such as pre-1940 ("heritage") status, carparking, and views of land and water
- The value that people place upon neighbourhood characteristics such as the presence of pre-1940 buildings.

Our analysis is intended to establish a set of standard values that can be used in assessing the trade-offs associated with planning regulations or alternative residential development typologies. Information on the hedonic structure of property prices can be used to analyse the welfare implications of land use regulations – for example, by identifying the marginal value of added residential floorspace in different areas of the city, or the value of other building or neighbourhood characteristics. Sheppard (1999) discusses some of the challenges and opportunities associated with using hedonic price models for welfare analysis.

However, it is difficult to robustly measure the degree to which specific planning regulations improve amenity or impose costs. In principle, observed property prices may be high due either to costs imposed by the planning system or amenities generated by it. We would expect regulatory constraints on housing supply to push up prices across the board, while amenities generated by the planning system will often, although not always, be more localised[3].

As a result, a hedonic price model estimated using property prices within Auckland will tend to capture the impact of localised amenities that drive differences between property prices in different areas. A different empirical strategy may be required to understand whether planning regulations are constraining housing supply.

[2] Or, more precisely, the sign and statistical significance of the regression coefficients on these explanatory variables.

[3] In other words, we would expect the effect of many amenities to be concentrated within neighbourhoods or to fall off relatively quickly with distance. For example, a house with attractive landscaping may have an impact on the value of houses on the same street but not houses in the neighbouring suburb.

# 2.0 Literature review

Hedonic price modelling has been used extensively to understand what property characteristics or attributes influence the amount people are willing to pay for those properties. In this section, we review some of the empirical literature on property prices.

## 2.1 Previous hedonic price studies in Auckland

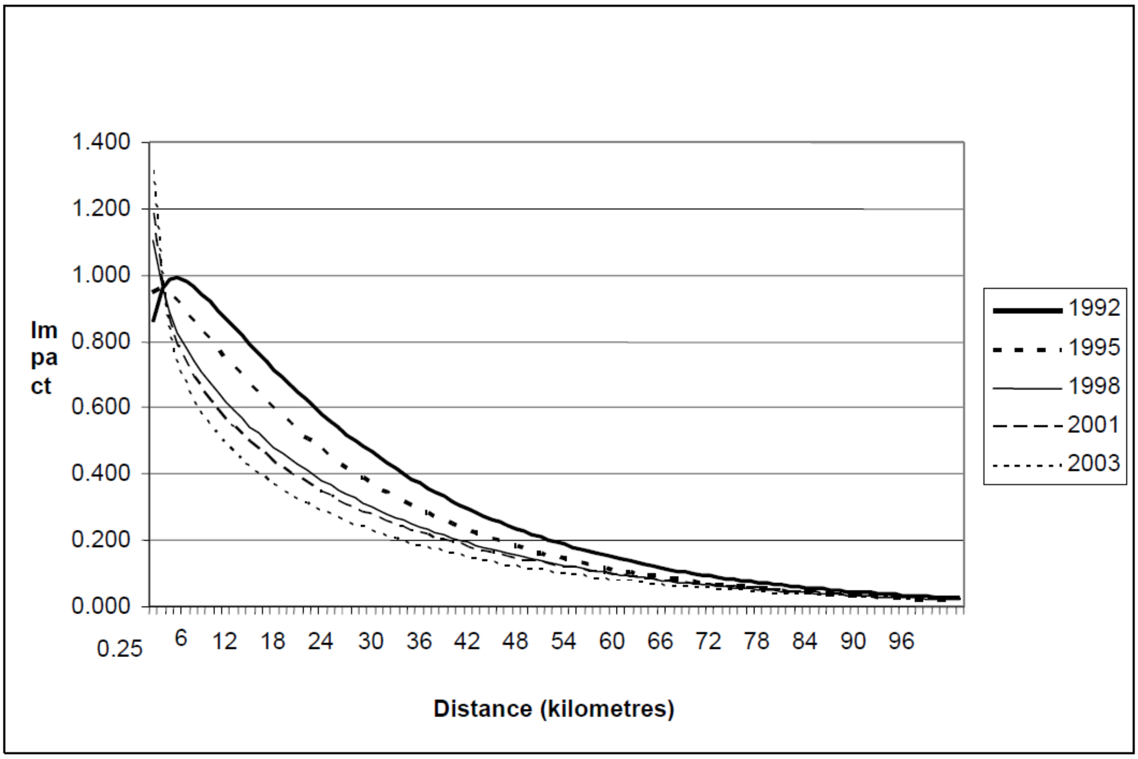
In Auckland, a number of studies have used data from the 1990s and early 2000s to understand how prices are affected by different property factors. These studies have revealed several “stylised facts” about property prices in Auckland. We have used these findings to motivate our identification of a preferred hedonic price model. However, we also test the impact of additional variables, including pre-1940 buildings within the neighbourhood:

- Larger lots and larger buildings are associated with higher sale prices – i.e. buyers value having more space. Interestingly, prices seem to be more responsive to building size (floorspace) than they are to land area (Bourassa, et al., 2003; Samarasinghe and Sharpe, 2010; MRCagney, 2013)
- Building quality features are associated with sale prices – including the age of buildings, exterior construction materials, and features such as decks or garages (Rehm, Filippova and Stone, 2006; Bourassa, et al., 2003; Rehm, 2009; Samarasinghe and Sharpe, 2010; MRCagney, 2013)
- Proximity to the city centre is associated with higher sale prices, as is proximity to the coast and school zoning (Grimes and Liang, 2007; Rohani, 2012)
- The Metropolitan Urban Limit (MUL) is associated with a “boundary discontinuity” in sale prices (Grimes and Liang, 2007, Zheng, 2013)
- Neighbourhood-level amenities, such as landscaping and views of water, are associated with higher property values (Bourassa, et al., 2003; Samarasinghe and Sharpe, 2010; Rohani, 2012; Filippova, 2009). School zoning also influences property values (Rohani, 2012; Rehm and Filippova, 2008).

The relationship between these attributes and property prices tended to be relatively consistent between different studies, with coefficients generally exhibiting the same sign and statistical significance. Some studies have found spatial variation in coefficients on land and floorspace variables (Donovan, 2011) and water views indicators (Filippova, 2009). Moreover, several studies find evidence that the magnitude of some coefficients has changed over time, which may be attributable to changes to the property market structure and prices over the last decade.

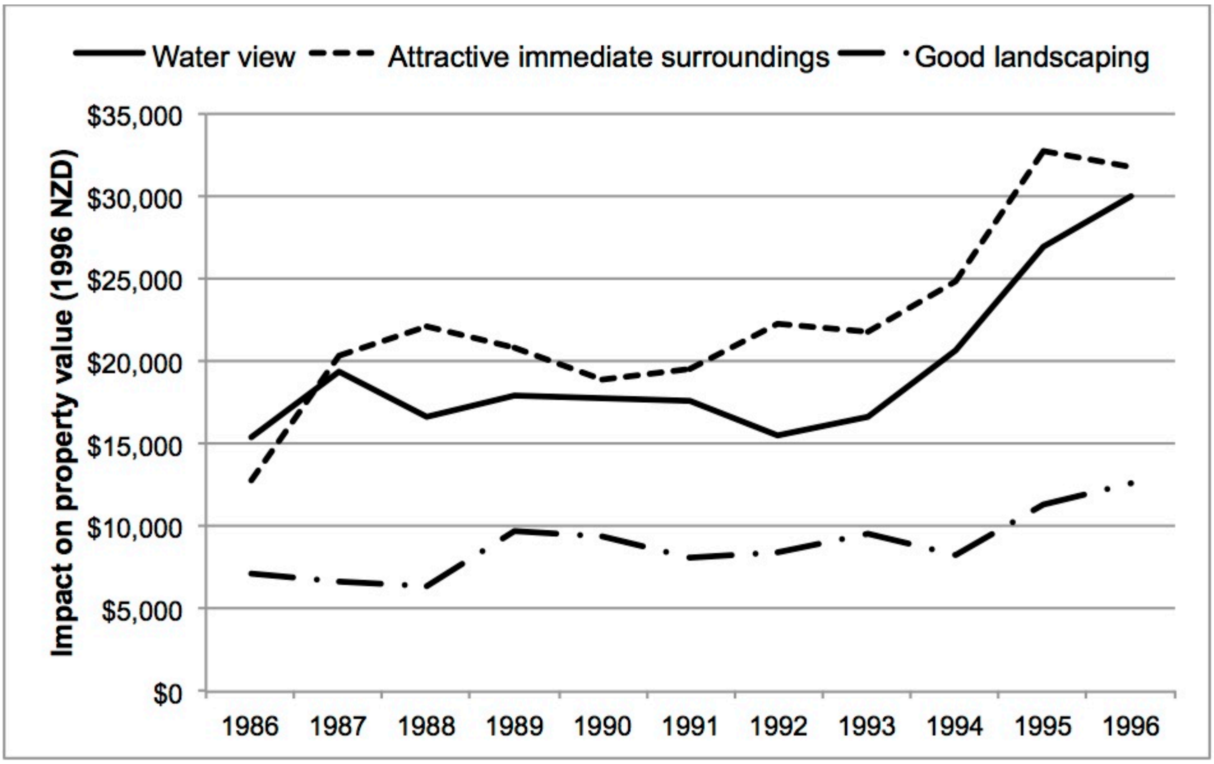
Grimes and Liang (2007) find that proximity to the city centre became increasingly valued over the period from 1992 to 2003. Their results are summarised in Figure 2. Interestingly, they find that between 1992 and 1998 close proximity to the city centre changed from being a disamenity (i.e. associated with lower land values) to an amenity (i.e. associated with higher land values).

Figure 2: Impact of distance from the CBD on real land values, 1992-2003 (Source: Grimes and Liang, 2007)




Similarly, Bourassa, et al. (2003) find that the hedonic value of aesthetic externalities increased more rapidly than house prices from 1986 to 1996. Their findings, which are summarised in Figure 3, suggest that the real hedonic value of water views rose by 97 per cent over the decade, the value of attractive immediate surroundings rose by 148 per cent, and the value of good landscaping rose by 75 per cent. By comparison, real house prices only rose by 58 per cent over the same period.

Figure 3: The value of aesthetic externalities in Auckland, 1986-1996 (Source: Bourassa, et al., 2003)



## 2.2 Some relevant international literature

International studies have also used hedonic pricing models to examine the impact of similar dwelling and neighbourhood characteristics. The recent literature has focused on trying to account for some of the statistical problems associated with hedonic modelling, in particular, sample selection bias and spatial autocorrelation.

Sample selection bias occurs when inferences are made about populations based on non-random sampling. In hedonic modelling this occurs when the sampling frame based on properties which have been sold, which only represent a proportion of the housing stock available at any given time. This means that if a non-random sample of the housing stock is used, it can lead to biased estimates on for the population (Hwang and Quigley, 2004). This issue is generally more problematic when investigating amenities or features that are likely to have known non-random effects, such as building developments around transport hubs. While we do not explicitly attempt to address sample selection bias in this study, we note that the property sale data we are using has extensive coverage of Auckland's urbanised area (see Figure 4).

Diao (2015) investigated this issue when examining the effects of subway stations on property prices in Boston, and accounted for this problem by applying a methodology based on the Heckman selection model, which can account for this sample selection problem. The study also utilised methods to account for spatial autocorrelation, which occurs when a property value in one location is dependent on values in neighbouring areas. Diao (2015) addresses this issue by applying spatial lag and error techniques, which provide more robust coefficients when compared to the baseline OLS hedonic model.

Lazrak, et al. (2014) used similar methods to understand the impact of heritage areas on property value in Zaanstad, a small city of approximately 150,000 people. This study used a baseline OLS hedonic model, but also tested spatial error and spatial lag models. It found a strong relationship between heritage attributes and property prices. It also found that prices were influenced by the size of properties and their proximity to water. In contrast to the Auckland results, the study found no a significant relationship between property prices and distance to the city centre. This may be due to Zaanstad's relative compactness, which reduces the relative benefit gained from locating in the centre is lower when compared to Auckland or Boston.

Studies have also extended these techniques to explore the interregional effects of different property attributes on house prices across Spain over time. McGreal and de La Paz (2013) used a complex two-stage least squares approach that incorporated instruments, and examined the causality of property attributes on house prices. This study reinforced the results of the Auckland level analysis and found that attributes such as building size and quality, views, access to transport and schooling tend to have a significant effect on house prices. It also found that over time, there was significant variation in the coefficients, indicating that the relative importance of these attributes is not static over time, which may be a function of people's changing preferences for different property attributes (McGreal and de La Paz, 2013)

Helbich, et al. (2014) conducted a similar interregional analysis in Austria and also found similar results to both the other international and Auckland level studies.

## 2.3 Choice of models

All the studies discussed in this section used an OLS hedonic model as the baseline hedonic model. Grimes and Liang (2007), Samarasinghe and Sharpe (2010) and various international studies then extended the analysis by applying techniques to account for spatial autoregression. Some studies also applied alternative model specifications, such as generalised linear models (GLM) or geographically weighted regressions (GWR) (Diao, 2015; Lazrak, et al., 2014; Helbich, et al., 2014). These approaches may be applicable when there is clear evidence that an assumption associated with OLS does not hold, such as when the when the population distribution is known to follow a non-normal distribution, or when the data exhibits strong heteroskedasticity (McGreal and de La Paz, 2013).

A limitation of using these alternative model specifications is that the models become more difficult to implement and the coefficients more difficult to interpret and understand (Hwang and Quigley, 2004). This is a potential reason why most of the Auckland-specific analysis tends to focus on improving the OLS model through spatial modelling. It also highlights that effective hedonic modelling is an iterative process that focuses on improving the robustness and accuracy of the modelling over time, rather than suggesting that any given model provides an exact value of a building attribute or property feature.

In this report, we follow the Auckland-specific literature, beginning with an OLS model and then testing for spatial dependence, which would result in a violation of OLS assumptions. We then test spatial error and spatial lag models to address any observed spatial dependence. In doing so, we note that some previous studies (Donovan, 2011; Filippova, 2009) have found evidence that model coefficients may vary between suburbs, which suggests that there may be a rationale to apply a GWR approach to the data. We chose not to do so at this stage due to the fact that outputs from a GWR model are less easily interpreted and thus not necessarily useful for informing a welfare analysis of urban policies.

# 3.0 Overview of data

Our analysis is primarily based on data from Auckland Council's property sales audit file, which includes information on all property transactions in Auckland. As a result, it presents a robust view on the current hedonic structure of the Auckland property market. We relied upon four key sources of information:

- An extract from the Auckland Council (2015) *District Valuation Roll* (DVR) database, which includes detailed information on property sales recorded in the region
- Geographic information system (GIS) analysis to identify the location of properties, their proximity to amenities such as the city centre (CBD) and the coast, and their proximity to other property sales
- Data from the 2013 Census to identify key socioeconomic characteristics of meshblocks, including median household income and population density (Statistics NZ, 2014).

As all of this data is inherently spatial, it was possible to relate property sales records to data organised by meshblocks. In this section, we describe this data and discuss how we compiled a dataset for analysis.

## 3.1 Overview of sales audit file

We obtained an extract from the DVR database maintained by Auckland Council (2015) that covers the years 2011 to 2014[4]. This database is maintained by the Council as an input into ratings valuations that are conducted every three years and compiled in accordance with ratings valuation rules published by Land Information New Zealand (2010). As a result, it contains data on the following attributes of properties (key model variables in parentheses):

- The location of the property – defined by both the corresponding Auckland Council rates valuation reference and the property's street address
- The date when the property was sold (SALEYEAR)
- The gross sale price including chattels (e.g. furniture and appliances) and the net sale price (NETPRICE)
- Land use data, including zoning, actual property use (e.g. type of residential or commercial property), number of units, and number of off-street carparks (CARPARKS)
- Data on the estimated decade of construction (used to identify PRE1940 status[5]) and condition of the primary buildings on the property (COND_WALL, COND_ROOF)

[4] Because this dataset was obtained in December 2014, it excludes some sales from the end of the 2014 calendar year.

[5] 1940 was used as the cutoff for heritage status for two reasons. First, it aligns with Auckland Council's built heritage policies, which aim to control the demolition of pre-1944 buildings. Second, previous research into the influence of vintage on Auckland property values has found evidence of a price premium for buildings constructed prior to the 1940s, but not after (Rehm, Filippova and Stone, 2006).

- Land area (LAND), building size (gross floor area) and site coverage
- Mass appraisal data, including the total living area (FLOORSPACE), presence of decks, workshops, garages, and the view from the building (VIEW).
- Other details of the sale, such as whether the property is sold freehold, leasehold, or in some other way.

In order to enable spatial analysis of this dataset, we associated each sales record with the longitude and latitude coordinates[6] of the relevant rating unit (as at 18 February 2015). Because data on rating units is continuously updated, with titles regularly being created and destroyed, it was not possible to match a small number of sales.

## 3.2 Additional variables joined to sales audit file

After geocoding the data, it was possible to join it to spatial data from other datasets, including meshblock-level from the Census and from other Auckland Council datasets.

We used GIS analysis to identify the 2013 Census meshblock associated with each sale record. Meshblocks were used for two reasons. First, this is the most fine-grained level at which Statistics New Zealand makes Census demographic data available. Second, meshblocks are likely to provide a reasonable representation of the "neighbourhood" around each property as they generally include around 20-80 residences that are bounded by roads, parks, or natural barriers (Torshizian and Grimes, 2014a).

We joined the following meshblock-level variables to each sale record:

- Straight-line distance from the meshblock centroid to the city centre (DCBD)[7] and to the coast (DCOAST)[8]
- Number of pre-1940 buildings within each meshblock (MBHERITAGE) – this variable was created using GIS analysis of Auckland Council's DVR database. This represents the best available estimate of the number of properties with potential heritage significance in Auckland.

---

[6] The coordinate reference system for these points is the New Zealand Transverse Mercator 2000 projection (EPSG:2193). This is standard practice for the majority of data published by New Zealand governments or about New Zealand.

[7] One alternative would be to use road network distance to the city centre, as travel distances are affected by Auckland's geography and infrastructure. This would result in a more realistic estimate of distances for some places but not for others. For example, although the road network distance from Devonport, a harbourside suburb on the North Shore, to the city centre is approximately 14 kilometres, actual travel distances are much smaller due to the presence of a frequent ferry service.

[8] Meshblock-level data was used here to reduce computational requirements. However, it is worth noting that this will tend to introduce some small spatial correlations between sales in the same meshblock. Torshizian (2014) presents a new package for the Stata statistical analysis programme that attempts to overcome these computational limits.

- Median household income within the meshblock, from the 2013 Census (HHINCOME). We did not adjust this value to account for wage inflation between 2011-2014. Due to the fact that the annual change in Statistics New Zealand's (2015) *Labour Cost Index* was consistently below 2 per cent during this period, we did not expect the differences from 2013 incomes to be significant.
- Population density (usually resident population per hectare) within the meshblock, from the 2013 Census (DENSITY).

Other meshblock-level variables were also available from previous RIMU analysis, including data on nearby transport infrastructure and other amenities such as public parks. In addition, it would also be possible to incorporate information on nearby schools – e.g. school decile ratings. Although these variables were not included in this model as they were not directly relevant to the aims of the study, it would be possible to incorporate them into a future study.

## 3.3 Filters applied to data

It was necessary to filter the data to exclude property sales that had missing variables or unsuitable variable values. For example, we excluded properties sold with no land or floorspace, as including those values would not enable us to use a logarithmic model specification[9].

We applied the following filters to the data:

- Exclude all sales records with missing or non-complying data in the variables tested for inclusion in the regression model[10]
- Exclude all sale records which had a zero value for the following variables: LAND, DCBD, DCOAST, HHINCOME, DENSITY[11]
- Exclude all non-residential property sales (i.e. only include sales with an actual property use value between 90 and 99)[12].
- Exclude all sale records with: FLOORSPACE less than 20 square metres and NETPRICE less than $10,000[13]

After applying these filters, we were left with a total of 72,855 usable sales records, out of a total of 142,449 sales records in Auckland Council's sales audit file.

[9] As discussed in the following section, we employed a logarithmic model specification to control for heteroskedasticity in the data. This is a common tactic in hedonic price studies.

[10] This resulted in the removal of approximately 21,300 sales.

[11] Zero values would prevent us from taking the logarithm of these values. The exclusion of properties sold with no land resulted in the removal of 42,200 sales, while other filters had a relatively minor effect. Some properties sold with no land represent sales of apartments or leasehold properties, while others may be data entry errors. In principle, GIS analysis could be used to correct for data entry errors.

[12] This resulted in the removal of approximately 5,400 data points.

[13] The aim of this is to prevent the inclusion of sales of outbuildings or garages mistakenly classified as residential sales. After applying the previous filters, this resulted in the removal of less than 300 sales.

## 3.4 Summary statistics of final dataset of residential property sales

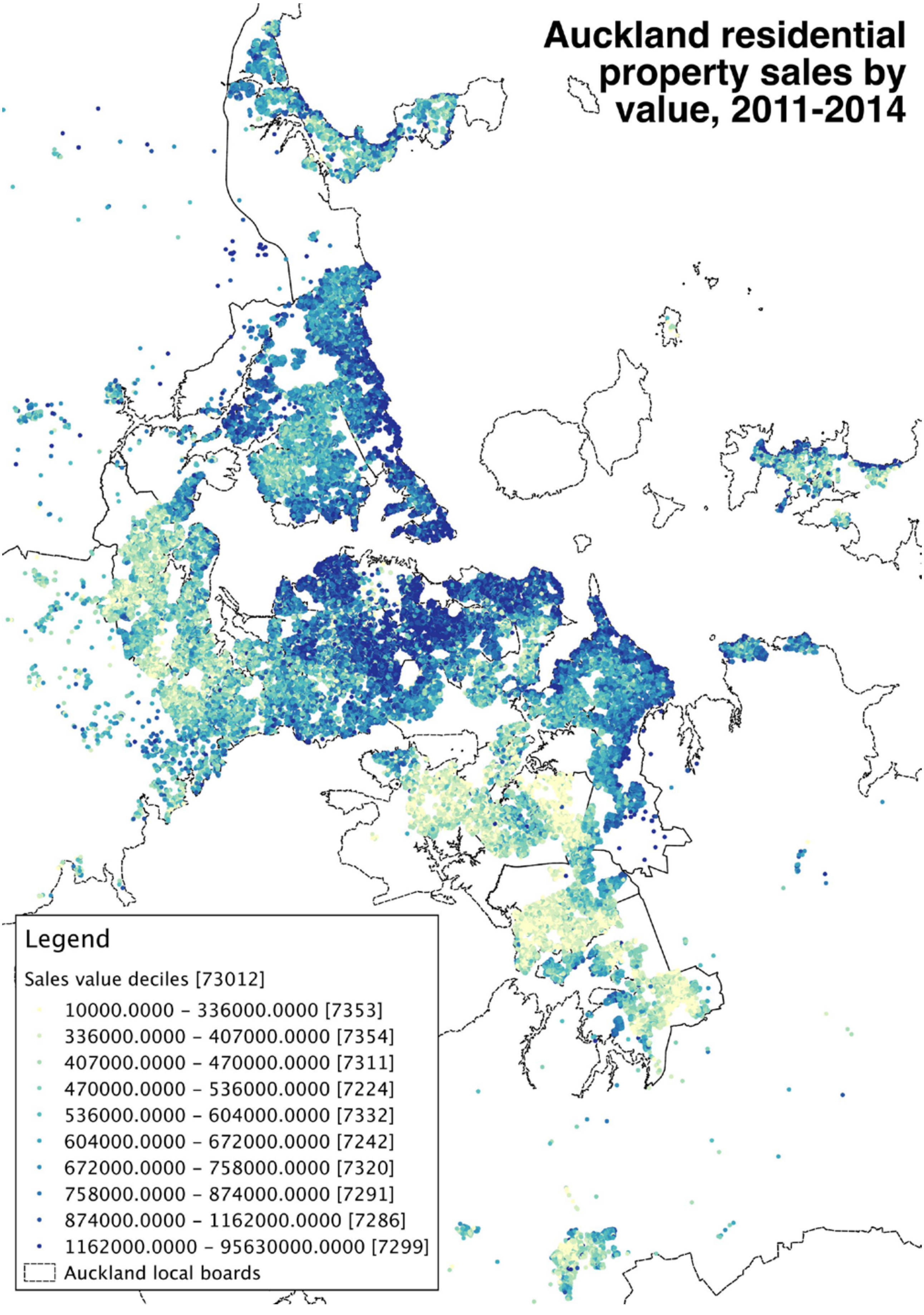
After joining datasets and filtering out unsuitable values, we obtained a final dataset of 72,855 residential property sales. Descriptive statistics about key variables are summarised in Table 2.

Table 2: Summary statistics of residential sales dataset, 2011-2014 (n=72,855)

| Statistic | Unit | Mean | Std Dev | Min | Max |
|---|---|---|---|---|---|
| X coordinates | NZTM2000 (metres) | 1,758,899 | 9,886 | 1,711,391 | 1,824,563 |
| Y coordinates | NZTM2000 (metres) | 5,917,523 | 15,328 | 5,874,246 | 5,996,054 |
| NETPRICE | Current NZD | $715,080 | $653,689 | $10,000 | $95,630,000 |
| LAND | Hectares | 0.15 | 12.76 | 0.001 | 3131.00 |
| FLOORSPACE | Square metres | 145 | 67 | 20 | 3,987 |
| DCBD | Metres | 15,009 | 10,845 | 147 | 101,187 |
| DCOAST | Metres | 1,290 | 1,404 | 2 | 9,983 |
| PRE1940 | Dummy | 15.3% | Not applicable | | |
| CARPARKS | Number | 1.6 | 3.0 | 0 | 302 |
| VIEW:NO VIEW | Dummy | 61.3% | Not applicable | | |
| VIEW:OTHER | Dummy | 25.5% | | | |
| VIEW:WATER | Dummy | 13.2% | | | |
| COND_ROOF:AVERAGE | Dummy | 31.5% | | | |
| COND_ROOF:FAIR | Dummy | 1.9% | | | |
| COND_ROOF:GOOD | Dummy | 65.9% | | | |
| COND_ROOF:POOR | Dummy | 0.5% | | | |
| COND_ROOF:MIXED | Dummy | 0.3% | | | |
| COND_WALL:AVERAGE | Dummy | 30.4% | | | |
| COND_WALL:FAIR | Dummy | 1.8% | | | |
| COND_WALL:GOOD | Dummy | 66.9% | | | |
| COND_WALL:POOR | Dummy | 0.6% | | | |
| COND_WALL:MIXED | Dummy | 0.3% | | | |
| MBHERITAGE | Number | 4.6 | 7.8 | 0 | 70 |
| HHINCOME | Current NZD | $86,507 | $29,381 | $2,500 | $150,000 |
| DENSITY | Residents / hectare | 32.9 | 18.0 | 0.01 | 806.3 |
| SALEYEAR:2011 | Dummy | 14.1% | Not applicable | | |
| SALEYEAR:2012 (3) | Dummy | 31.8% | | | |
| SALEYEAR:2013 (3) | Dummy | 32.4% | | | |
| SALEYEAR:2014 (3) | Dummy | 21.7% | | | |

Figure 4 maps these residential property sales, coloured by decile of value (dark blue = high value; yellow = low value). It illustrates some broad features of the Auckland property market, including the concentration of high-value properties in inner isthmus suburbs, coastal areas of the North Shore, and the eastern bays, and the concentration of low-value properties in south Auckland and, to a lesser extent, west Auckland. Note that our data does not include many sales in the Auckland city centre, as city centre residential properties tended to be apartments or leasehold properties sold with no land. Likewise, no residential sales were recorded in the industrial areas of Mount Wellington-Penrose, East Tamaki, the Auckland Airport, and Albany industrial area, or in Albany centre.

Figure 4: Auckland residential property sales, 2011-2014, coloured by decile of value



# 4.0 Empirical strategy

Here, we describe our approach to developing a regression model that predicts residential property sales as a function of dwelling characteristics, dwelling location, and neighbourhood characteristics.

Our empirical strategy was as follows:

- First, we tested ordinary least squares (OLS) models, finding a preferred model which had maximum predictive power for residential property sale prices. This analysis is described in Section 4.1.
- Second, we tested the preferred OLS model for heteroskedasticity (using the Breusch-Pagan test) and spatial dependence (using Moran's I), finding that (a) the OLS model exhibited heteroskedasticity and (b) model residuals exhibited spatial dependence. This analysis is discussed in Section 4.2.
- Third, we tested four alternative spatial regression models to attempt to control for spatial dependence. Due to the fact that spatial regression models are computationally intensive, it was necessary to estimate these models on a randomly selected subset of the data. As discussed in Section 4.3, we tested spatial error and spatial lag models that employed two alternative definitions of "neighbouring" properties.
- Fourth, we identified a preferred spatial regression model based on a comparison of AIC scores and an analysis of remaining spatial dependence in model residuals. As discussed in Section 4.4, we found that a spatial error model with a 1km radius neighbourhood was the preferred model. This is consistent with previous research on property prices in Auckland (Grimes and Liang, 2007).

All analysis was conducted in R, using the "car", "lmtest", "sp" and "spdep" packages.

## 4.1 Identification of preferred OLS model

First, we identified a preferred OLS regression model by testing several alternative models of the form:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Where NETPRICE is the dependent variable ($y_i$, an nx1 vector), $\alpha$ is a constant term to be estimated, $x_i$ is a nxk vector of k explanatory variables (including variables of interest such as LAND, DCBD, and PRE1940, control variables such as HHINCOME and DENSITY, and time dummies), $\beta$ is a 1xk vector of coefficients to be estimated, and $\varepsilon_i$ is an nx1 vector of error terms.

We conducted the following analysis (in R) to identify a preferred model:

- First, we started by estimating a relatively expansive OLS model that included all of the thirteen variables identified in Table 2 above.

- Second, we re-estimated the OLS model after step-wise removal of variables with Variance Inflation Factor (VIF) scores greater than 3[14]. This resulted in the removal of the COND_ROOF variable, which was found to be highly collinear with the COND_WALL variable. It was not necessary to remove any other variables – all were found to have a VIF score under 2. We also estimated several other OLS models, including one which also excluded the COND_WALL variable, and one which removed the demographic measures from the 2013 Census, for completeness.
- Third, in order to choose a preferred OLS model from this set, we estimated Akaike's Information Criterion (AIC) scores for each of the four models[15]. We found that the second model, which included all variables except COND_WALL, had the best AIC score. This indicates that it offers the best mix of goodness of fit ($R^2$) and degrees of freedom (df).

Table 4 in Appendix A summarises the OLS models that we tested, with our preferred OLS model highlighted in yellow. The overall model is highly statistically significant (p-value<0.01) and explains approximately 64 per cent of the variation in sale prices (Adjusted $R^2$=0.64)[16].

In addition, all of the key variables are highly statistically significant (p-value<0.01) and have the expected sign. For example, this model suggests that more floorspace is associated with higher sale prices, views of water are associated with higher sale prices, and proximity to the city centre is associated with higher sale prices. This is reassuring as it suggests that our intuitions about home-buyers' preferences are reasonable.

## 4.2 Investigating spatial dependence in OLS model residuals

However, this analysis is not sufficient to establish whether our preferred OLS model is efficient. Regression analysis requires errors to be uncorrelated – i.e. for there to be no patterns in model residuals. Unobserved spatial processes, including correlations between the value of neighbouring properties or un-measured characteristics of neighbourhoods, can violate this assumption of independence.

Mapping and inspection of residential sales values in Figure 4 above suggests there may be some unobserved spatial processes at work. This intuition is reinforced by tests for heteroskedasticity

[14] VIF scores indicate whether there is any multicollinearity between the explanatory variables in the model. Multicollinearity implies that two (or more) explanatory variables may in fact be measuring the same phenomenon. Failing to remove collinear variables is likely to result in an inaccurate estimate of regression coefficients. VIF scores are calculated by running a series of OLS regressions on the explanatory variables that attempt to predict each explanatory variable as a linear combination of all other explanatory variables. The VIF score for each variable is then calculated as a function of the goodness of fit of the OLS model – 1 / (1- $R^2$). Therefore, a VIF score of 3 would indicate that 66 per cent of the variation in a single explanatory variable could be explained as a function of the other explanatory variables. This is a common threshold to use when evaluating multicollinearity.

[15] Akaike's Information Criterion measures the trade-off between the model's goodness of fit ($R^2$) and the degrees of freedom in the model.

[16] This $R^2$ value is low compared to the results from other studies of Auckland property values, which tend to be in the range of 0.7-0.8. This may be worth further investigation.

(non-constant variance) in model residuals. We applied the Breusch-Pagan test for heteroskedasticity in linear models to our preferred OLS model, finding that residuals are systematically biased[17]. This suggests that we have omitted some relevant information from the model.

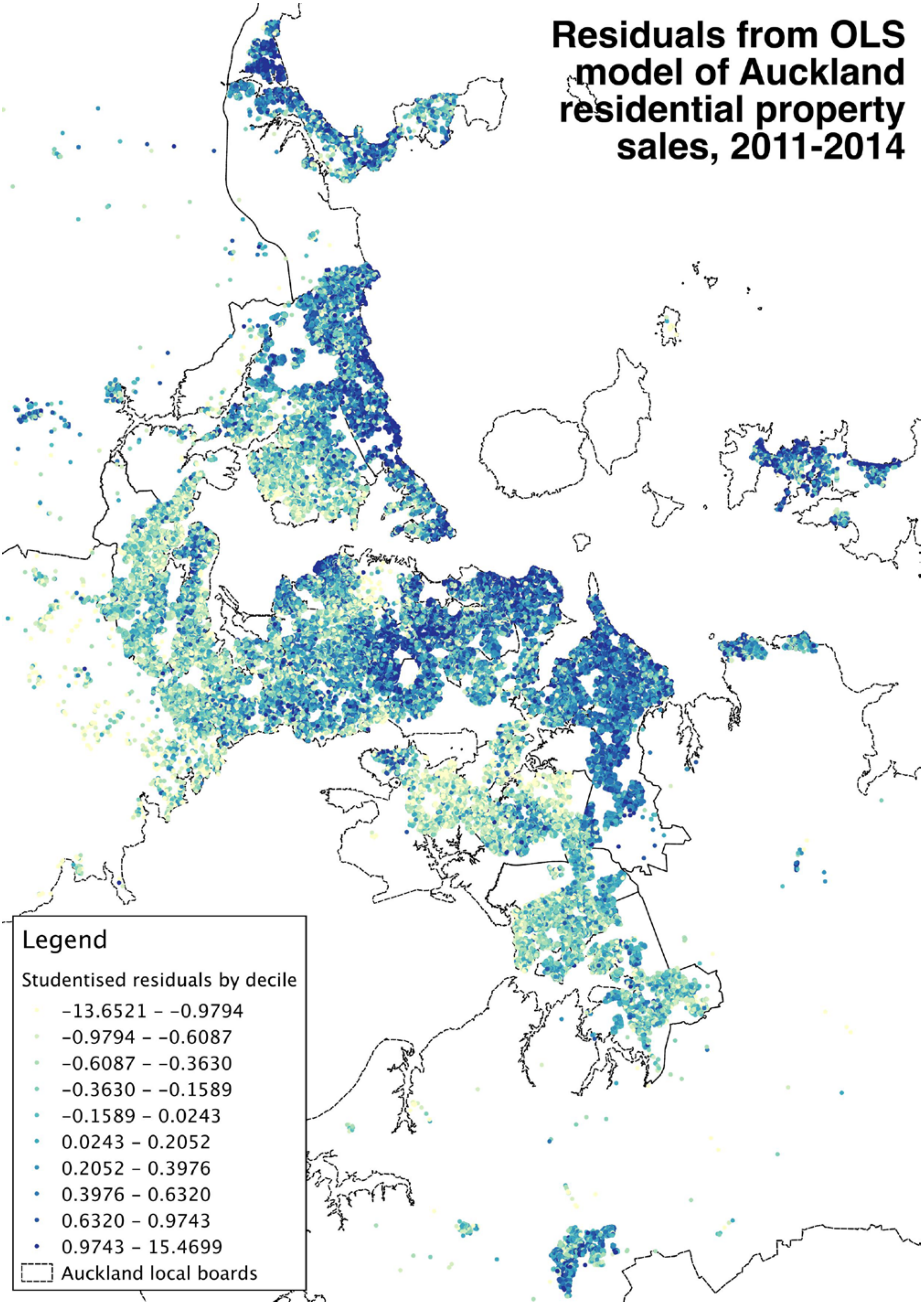
Figure 5 maps residuals from the OLS model. It shows that there is a strong spatial pattern in the residuals – positive and negative errors tend to be clustered near each other. This indicates that we cannot assume that errors are uncorrelated with each other. We confirmed this intuition by using Moran's I to test spatial autocorrelation between nearby model residuals[18]. We found a Moran's I statistic standard deviation of 85.09 and a highly statistically significant p-value (< 2.2e-16). Effectively, the amount of spatial autocorrelation in the model residuals is too large to be explained by chance.

As a result, it is necessary to investigate spatial regression models to control for the spatial processes underlying property sales.

[17] BP score = 1019.281; df=18; p-value < 2.2e-16.

[18] Due to computational requirements, we applied Moran's I to a randomly selected subset of 10,000 sales. We defined "neighbouring properties" as any sales records within a one-kilometre radius. Our rationale for and approach to taking a subset of the data are discussed in more detail in the Section 4.3.1.

Figure 5: OLS model residuals, by decile



## 4.3 Specifying spatial regression models

There are two main approaches to spatial regression that treat the spatial processes underlying the data in slightly different ways:

- Some types of models treat spatial processes as a "nuisance" to be eliminated or controlled. This is the approach underpinning a spatial error model.
- Other types of models treat spatial processes as a substantive effect of interest. They build spatial relationships into the model as parameters to be estimated. This approach underpins a spatial lag model (as well as other types of models such as geographically weighted regressions).

In this paper, we test both approaches and identify a preferred model based on Akaike's Information Criterion and an analysis of remaining spatial autocorrelation in model residuals.

### 4.3.1 Taking a random subset of the data

Spatial regression imposes significant computational requirements due to the fact that it is necessary to analyse the relationships between a large number of points. In this case, we found that it was not feasible to estimate spatial regression models on the full Auckland property sales dataset, which included 72,855 points.

We addressed this by randomly selecting a subset of 10,000 points from the full dataset[19]. We ran the preferred OLS regression model on the subset, finding that the results closely matched our OLS analysis of the full dataset. A comparison of the two models is presented in Table 5 in Appendix A.

### 4.3.2 Defining a neighbourhood

There are multiple ways to define a "neighbourhood" for purposes of spatial analysis (Harris, 2013). In this paper, we consider two definitions of a property's neighbours for purposes of analysis:

- Sale records located within a one-kilometre radius (K1KM)
- Sale records located within the same meshblock (KMB).

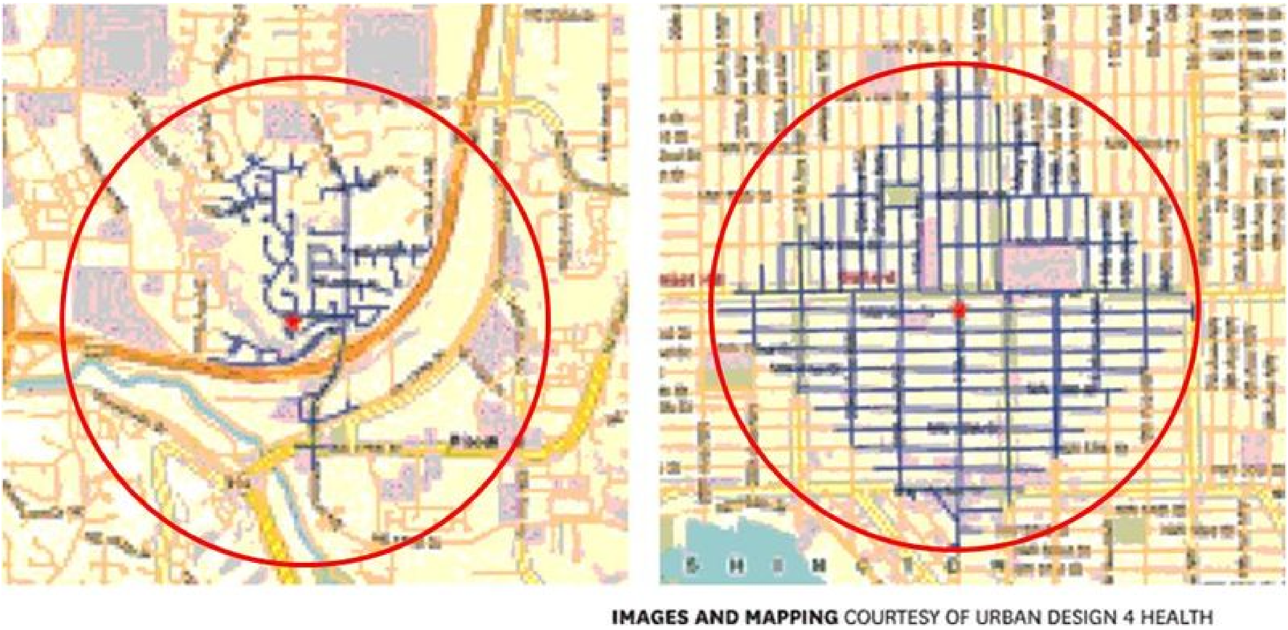
Previous research into the relationship between dwelling and neighbourhood features and residential satisfaction in Auckland has identified these as the most relevant definitions of a residential neighbourhood (2014a). In particular, Torshizian and Grimes (2014b) find that a "dynamic" definition of a neighbourhood, which uses road network analysis to identify properties located within a 15-minute walking distance, performs best.

Here, we have used a one-kilometre radius around to approximate 15-minute walking distance from residential properties. We have assumed that people walk, on average, one kilometre every

[19] For the sake of replicability, we provided an arbitrarily chosen seed number to R's random number generator. Seed number and other code is available on request.

twelve minutes, and that the structure of street networks will tend to mean that people cannot simply walk in a straight line. As illustrated in Figure 6, a one-kilometre radius may overestimate walking catchments in some areas and underestimate them in others.

Figure 6: Walking catchments vary depending upon street networks and natural barriers (Source: Wieckowski, 2010)




We used spatial analysis tools in R to identify neighbouring properties using each of these definitions[20]. We then created two spatial weights matrices with row-standardised weights for each row[21].

### 4.3.3 Spatial error models

A spatial error model examines spatial dependence between the residuals, or error terms, of neighbouring data points. Essentially, it treats spatial autocorrelation as a "nuisance" that the model must control, rather than a meaningful process that it is attempting to explain.

In order to do so, spatial error models decompose the error into two parts: a spatially autocorrelated component and a remaining uncorrelated component. They take on the following form:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

[20] We identified neighbours within a 1km radius using the "dnearneigh" function in the "spdep" package. It was necessary to write a simple function to identify neighbours within meshblocks – code available on request.

[21] Spatial weights matrices were created using the "nb2listw" function in the spdep package. Row-standardisation means that if a data point has *n* neighbours, each will be assigned a weight of *1/n* for analysis. Therefore, a point with five neighbours would have weights of 0.2, while a point with two neighbours would have weights of 0.5. Points with no neighbours were dropped from the analysis.

Where NETPRICE is the dependent variable ($y_i$), $\alpha$ is a constant term to be estimated, $\beta$ is a vector of coefficients to be estimated, $x_i$ is a vector of explanatory variables (including variables of interest such as LAND, DCBD, and PRE1940, control variables such as HHINCOME and DENSITY, and time dummies), and $\varepsilon_i$ is an error term.

The error term is in turn decomposed into two parts as follows:

$$\varepsilon_i = \lambda W_{ij}\varepsilon_j + \xi_i$$

Where $\varepsilon_j$ is a vector of error terms for $j \neq i$, weighted using spatial weights matrix $W_{ij}$ (based on either K1KM or KMB), $\lambda$ is the spatial error coefficient, and $\xi_i$ is a vector of uncorrelated error terms. We estimated spatial error regression models on the Auckland residential sales dataset using a spatial analysis package in R ("spdep")[22].

### 4.3.4 Spatial lag models

By contrast, a spatial lag model treats spatial dependence as a process of interest that the model seeks to explain. It attempts to explain the value of a data point partly in terms of the characteristics of neighbouring data points. For example, this may mean modelling the sale price of a single house as a function of the sale price of neighbouring properties. (Or, equally, of other characteristics of neighbouring properties, such as building size or condition.)

In order to do so, spatial lag models incorporate a "spatially lagged" variable on the right hand side of the regression equation. They take on the following form:

$$y_i = \alpha + \beta x_i + \rho W_{ij} y_j + \varepsilon_i$$

Where NETPRICE is the dependent variable ($y_i$), $\alpha$ is a constant term to be estimated, $\beta$ is a vector of coefficients to be estimated, $x_i$ is a vector of explanatory variables (including variables of interest such as LAND, DCBD, and PRE1940, control variables such as HHINCOME and DENSITY, and time dummies), $y_j$ are dependent variables for $j \neq i$, weighted using spatial weights matrix $W_{ij}$, $\rho$ is the spatial coefficient, and $\varepsilon_i$ is an error term.

We estimated spatial lag regression models on the Auckland residential sales dataset using a spatial analysis package in R ("spdep")[23].

## 4.4 Identifying a preferred spatial regression model

We estimated four spatial regression models in total: two spatial error models estimated using the two alternative definitions of neighbouring properties (K1KM and KMB) and two spatial error models estimated using K1KM and KMB.

Table 3 presents a comparison of our preferred OLS model with the four spatial regression models we tested. We used AIC scores, which measure the combination of goodness of fit and degrees of freedom (df) offered by each model to select a preferred spatial regression model. We found that

[22] In particular, the errorsarlm function.

[23] In particular, the lagsarlm function.

the spatial error model with a 1km radius neighbourhood (highlighted in yellow) was the preferred model.

Table 3: OLS and spatial regression results for a randomly-selected subset of 10,000 residential sales

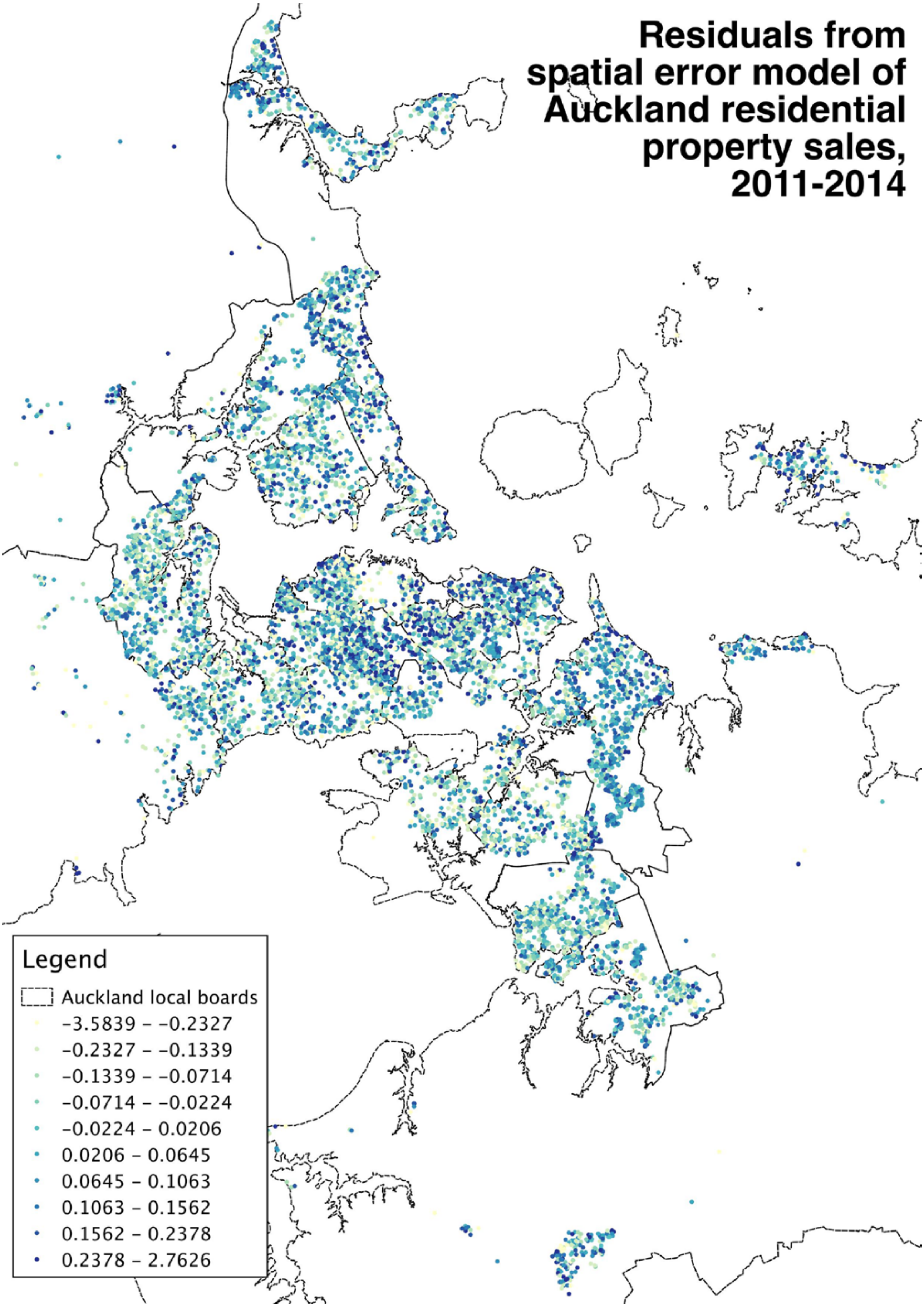
**Auckland residential property sales 2011-2014, OLS and spatial regression models**

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | log(sale_price_net) | | | | |
| | | Spatial error models | | Spatial lag models | |
| | OLS model | 1km radius neighbourhood | Meshblock neighbourhood | 1km radius neighbourhood | Meshblock neighbourhood |
| log(LAND) | 0.120*** | 0.209*** | 0.145*** | 0.139*** | 0.123*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| log(FLOORSPACE) | 0.609*** | 0.479*** | 0.570*** | 0.586*** | 0.607*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| log(DCBD) | -0.284*** | -0.340*** | -0.289*** | -0.278*** | -0.285*** |
| | (0.006) | (0.018) | (0.007) | (0.006) | (0.006) |
| log(DCOAST) | -0.014*** | -0.028*** | -0.014*** | -0.010*** | -0.014*** |
| | (0.003) | (0.005) | (0.003) | (0.003) | (0.003) |
| PRE1940 | 0.121*** | 0.098*** | 0.120*** | 0.119*** | 0.123*** |
| | (0.011) | (0.01) | (0.011) | (0.011) | (0.011) |
| CARPARKS | 0.001 | 0.0003 | 0.001 | 0.001 | 0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| VIEW:OTHER (1) | 0.006 | -0.017** | 0.0004 | 0.003 | 0.006 |
| | (0.008) | (0.007) | (0.008) | (0.008) | (0.008) |
| VIEW:WATER (1) | 0.164*** | 0.083*** | 0.141*** | 0.151*** | 0.164*** |
| | (0.011) | (0.01) | (0.011) | (0.01) | (0.011) |
| COND_WALL:FAIR (2) | -0.022 | -0.018 | -0.042** | -0.022 | -0.023 |
| | (0.022) | (0.019) | (0.021) | (0.022) | (0.022) |
| COND_WALL:GOOD (2) | 0.066*** | 0.064*** | 0.065*** | 0.068*** | 0.066*** |
| | (0.008) | (0.007) | (0.008) | (0.008) | (0.008) |
| COND_WALL:POOR (2) | -0.168*** | -0.086** | -0.164*** | -0.161*** | -0.168*** |
| | (0.041) | (0.036) | (0.04) | (0.041) | (0.041) |
| COND_WALL:MIXED (2) | -0.006 | -0.019 | -0.006 | 0.009 | -0.003 |
| | (0.053) | (0.045) | (0.049) | (0.052) | (0.053) |
| MBHERITAGE | 0.004*** | 0.003*** | 0.004*** | 0.004*** | 0.004*** |
| | (0.0005) | (0.001) | (0.001) | (0.0005) | (0.0005) |
| log(HHINCOME) | 0.194*** | 0.083*** | 0.208*** | 0.180*** | 0.191*** |
| | (0.01) | (0.01) | (0.011) | (0.01) | (0.01) |
| log(DENSITY) | -0.034*** | -0.011** | -0.027*** | -0.053*** | -0.033*** |
| | (0.004) | (0.005) | (0.005) | (0.004) | (0.004) |
| SALEYEAR:2012 (3) | 0.068*** | 0.101*** | 0.076*** | 0.070*** | 0.068*** |
| | (0.01) | (0.009) | (0.01) | (0.01) | (0.01) |
| SALEYEAR:2013 (3) | 0.197*** | 0.229*** | 0.205*** | 0.198*** | 0.197*** |
| | (0.01) | (0.009) | (0.01) | (0.01) | (0.01) |
| SALEYEAR:2014 (3) | 0.300*** | 0.327*** | 0.311*** | 0.302*** | 0.300*** |
| | (0.011) | (0.01) | (0.01) | (0.011) | (0.011) |
| Constant | 11.093*** | 13.756*** | 11.216*** | 10.661*** | 11.136*** |
| | (0.144) | (0.225) | (0.161) | (0.145) | (0.144) |
| Observations | 10,000 | 10,000 | 10,000 | 10,000 | 10,000 |
| Adjusted R2 | 0.625 | | | | |
| F Statistic | 927.097*** (df = 18; 9981) | | | | |
| Log Likelihood | | -1,540.51 | -2,354.63 | -2,585.20 | -2,720.93 |
| sigma$^2$ | | 0.077 | 0.091 | 0.098 | 0.101 |
| Akaike Inf. Crit. | 5,492.39 | 3,123.01 | 4,751.25 | 5,212.39 | 5,483.86 |
| Wald Test (df = 1) | | 5,108.288*** | 854.388*** | 281.183*** | 10.550*** |
| LR Test (df = 1) | | 2,371.383*** | 743.139*** | 281.998*** | 10.532*** |

*Notes:* *p<0.1; **p<0.05; ***p<0.01
(1) VIEW is a categorical (dummy) variable. The base level is "no view".
(2) COND_WALL is a categorical (dummy) variable. The base level is "average" condition.
(3) SALEYEAR is a categorical (dummy) variable. The base level is 2011.

Finally, we considered whether there is any remaining spatial autocorrelation in model residuals. Figure 7 maps the residuals from the preferred spatial error model, showing that error terms from this model lack the pronounced spatial pattern of the residuals from the preferred OLS model. However, Moran's I shows that there is still some remaining spatial dependence in the spatial lag model with a 1km radius neighbourhood (standard deviation of 77.5529 and a highly statistically significant p-value<2.2e-16). This tends to reinforce our preference for the the spatial error model.

Figure 7: Residuals from spatial error model with a 1km radius neighbourhood, grouped by decile



## 4.5 Key findings from hedonic regression analysis

In summary, we find evidence of spatial autocorrelation in both residential sale prices and OLS model residuals. As a result, we find that a spatial error model estimated using a 1km radius neighbourhood is the preferred model. This finding is consistent with earlier research on the Auckland housing market (Grimes and Liang, 2007, which employed a spatial error model to analyse property values) and residential satisfaction in Auckland (Torshizian and Grimes, 2014a, which identified a 15 minute walking catchment as the preferred definition of a neighbourhood).

However, a comparison of coefficients from the preferred OLS model and the spatial error model suggests that the spatial error model does not affect their sign or statistical significance. This comparison is reported in Table 3 (above). This suggests that although our OLS model of Auckland residential sale prices may produce biased estimates, the resulting differences in coefficients are not “economically” significant as their overall direction does not change.

# 5.0 Discussion and conclusions

Lastly, we discuss some implications of our analysis for policymakers and researchers.

## 5.1 Implications for policymakers

This report establishes some general facts about the hedonic structure of housing prices in Auckland. As discussed in the introduction, we have used higher property values as an indicator of amenity for households. In other words, if a characteristic of a property or neighbourhood is associated with a higher sale price, it suggests that people place a value on it.

Our first, most important finding is that <u>buyers exhibit a strong preference for more floorspace</u>. Based on the results from our preferred spatial error model, we find that properties with more floorspace or more land commanded higher prices. We note that the coefficient on the floorspace variable is considerably higher than the coefficient on the land variable in both the baseline OLS model and our preferred spatial error model estimated on a subset of property sales. While this does not provide a sufficient basis for firm conclusions, it is consistent with the results of earlier research (Bourassa, et al, 2003; Donovan, 2011). It also tends to strengthen the intuition – drawn from a casual observation of recent housing development outcomes – that Auckland's underlying challenge is not a scarcity of developable land, but a scarcity of floorspace. As a result, we would expect policies that enabled an increase in residential floorspace, either by enabling higher-density development or an increase in land supply for new subdivisions, to improve amenity for Aucklanders.

A second important finding is that <u>sale prices are influenced by location within the city</u>. People are not indifferent between different locations – all else being equal, they show a distinct preference to be closer to the city centre and a weaker, but still significant, preference to be close to the coast. Our analysis suggests that increasing distance from the city centre is associated with lower property values, as is increasing distance from the coast. This suggests that enabling increased housing supply in desirable locations will result in improved amenity relative to supplying new housing in less desirable locations.

Third, we find that other characteristics of dwellings are associated with higher sales prices. Even after controlling for some neighbourhood and location characteristics, people place a higher value on older (pre-1940) buildings. And, unsurprisingly, people prefer dwellings that are in good condition – houses with walls that are in good condition tend to sell for higher prices than houses with walls in average condition, while houses with walls in poor condition are worth less.

However, <u>carparking does not appear to have a strong impact on sales prices</u>. While the coefficient on the carparking variable was positive and statistically significant in the baseline OLS model, it is not statistically significant in the spatial error model estimated on a subset of property sales. In other words, the value of carparking may be quite marginal, and potentially lower than the cost to construct a single parking space, which ranges from $1,900-$2,200 for surface parking to $40,600-$46,100 for underground parking (Rawlinsons, 2014). This strengthens the findings of previous Auckland-specific research which has shown that minimum carparking requirements impose an opportunity cost on property owners (MRCagney 2013).

Fourth, our findings provide empirical evidence for the view that property markets are influenced by "highly localised externalities" related to neighbouring land uses. While we were not able to measure all potential neighbourhood effects, we did find evidence that proximity to pre-1940 buildings was associated with higher sales prices for both old and new buildings. Likewise, dwellings with a view of water commanded higher prices than comparable properties with no views. This suggests that policies that preserve these neighbourhood features may also preserve amenity for residents.

Lastly, we find evidence of spatial dependence in Auckland's housing market. In other words, the sale price of a single house is correlated with neighbouring property values. The OLS regression models that we tested could not fully explain these localised correlations, possibly due to omitted variables that we were not able to observe. We tested several spatial regression models, finding that a spatial error model (which treats spatial dependence as a "nuisance" to control) performed better than a spatial lag model (which treats spatial dependence as a process of interest to explain). This suggests that there are some unexplained spatial processes that influence property prices. However, the differences between our preferred spatial error model and an OLS model were not "economically" significant, as the sign and statistical significance of most coefficients remains the same.

## 5.2 Implications for researchers

Our findings are potentially relevant for policy analysis, including as an input into analysis of the welfare implications of planning regulations that either constrain the supply of floorspace (a valued amenity) or provide or protect other neighbourhood-level amenities (e.g. pre-1940 buildings). Sheppard (1999) provides some relevant guidance on the application of hedonic price modelling for welfare analysis.

There is an opportunity to do some further work in three key areas:

First, it would be useful to extend the model by including additional neighbourhood-level or meshblock-level explanatory variables and control variables, such as:

- Infrastructure variables, such as proximity to major roads, rapid transit infrastructure, accessibility to employment via public transport, the connectedness of street grids, etc.
- Measures of land use regulations, such as residential and commercial zoning or application of special purpose overlays (e.g. volcanic viewshafts)
- Proximity to other amenities, such as public parks or street trees.

Second, it would be useful to apply a similar methodology to study Auckland's commercial and industrial property markets. While Auckland Council's sales audit file contains fewer property sales in these categories, it is still a rich source of data on businesses' location preferences.

Third, our findings around the preferred model (a spatial error model with a 1km radius neighbourhood) and the economic significance of spatial error model coefficients relative to the OLS model coefficients are consistent with previous work. However, further work is needed to refine our estimates of the effects of dwelling and neighbourhood characteristics. This may include applying spatial regression modelling techniques to the full dataset, or using repeated sampling ("bootstrapping") to refine our estimates of regression coefficients and variances. Given previous findings from the literature, there may also be a case to test alternative spatial regression models, such as geographically weighted regression.

# 6.0 References


Anas, A., Arnott, R and Small, K. A. 1998. Urban Spatial Structure, *Journal of Economic Literature*, Vol. 36, No. 3 (Sep., 1998), pp. 1426-1464.

Auckland Council (2015), *Auckland property sales audit file*. Auckland: Auckland Council. Some information available online at http://www.aucklandcouncil.govt.nz/EN/ratesbuildingproperty/propertyinformation/Pages/PropertyFilesReportshome.aspx.

Chung, L.L.W. 1994. The economics of land-use zoning: A literature review and analysis of the work of Coase. *The Town Planning Review*, Vol. 65, No. 1 (Jan., 1994), pp. 77-98.

Bivand, R. S., Hauke, J and Kossowski, T. 2013. "Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods." *Geographical Analysis*, 45(2), 150-179.

Bivand, R.S., Keitt, T and Rowlingson, B. 2014. "rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-1." URL http://CRAN.R-project.org/package=rgdal.

Bivand, R.S., Pebesma, E.J and Gomez-Rubio, V. 2013. *Applied spatial data analysis with R, Second edition*. Springer, NY. URL http://www.asdar-book.org/.

Bivand, R.S and Piras, G. 2015. "Comparing Implementations of Estimation Methods for Spatial Econometrics." *Journal of Statistical Software*, 63(18), 1-36. URL http://www.jstatsoft.org/v63/i18/.

Bourassa, S., Hoesli, M and Sun, J. 2003. The price of aesthetic externalities, Geneva: FAME research paper no. 98.

Diao, M. 2015. Selectivity, spatial autocorrelation and the valuation on transit accessibility. Urban studies, 52(1), pp. 159-177.

Donovan, S. 2011. *Space Matters: Agglomeration Economies and Spatial Competition in New Zealand*. Amsterdam: VU University (Unpublished masters thesis).

Filippova, O. 2009. "The influence of submarkets on water view house price premiums in New Zealand." *International Journal of Housing Markets and Analysis* 2(1).

Fox, J and Weisberg, S. 2011. *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage. URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Glaeser, E. 2008. *Cities, Agglomeration, and Spatial Equilibrium*: The Lindahl Lectures. Oxford, Oxford University Press.

Glaeser, E.L., Gyourko, J and Saks, R.E. 2005. "Why is Manhattan so expensive? Regulation and the rise in house prices". Cambridge, NBER Working Paper Series, No. 10124.

Grimes, A., Apatov, E., Lutchman, L. and Robinson, A. 2014. "Infrastructure's long-lived impact on urban development: Theory and empirics." *Motu Working Paper* 14-11. Wellington: Motu Economic and Public Policy Research.

Grimes, A and Liang, Y. 2007. "Spatial Determinants of Land Prices in Auckland: Does the metropolitan urban limit have an effect?" *Motu Working* paper 07-09. Wellington: Motu Economic and Public Policy Research.

Grimes, A and Mitchell, I. 2015. "Impacts of Planning Rules, Regulations, Uncertainty and Delay on Residential Property Development," *Motu Working Paper* 15-02. Wellington: Motu Economic and Public Policy Research.

Harris, Richard 2013. *An Introduction to Mapping and Spatial Modelling R*. Bristol, UK: University of Bristol. URL: http://www.social-statistics.org.

Helbich, M., Brunauer, W., Vaz, E and Nijkamp, P. 2014. Spatial heterogeneity in hedonic house price models: The case of Austria. Urban studies, 51(2), pp. 390-411.

Hlavac, Marek. 2013. “stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables”. R package version 4.5.3. URL http://CRAN.R-project.org/package=stargazer.

Hwang, M and Quigley, J. 2004. Selectivity, quality adjustment and mean reversion in the measurement of house values. Journal of real estate finance and economics, 28(2), pp. 191-214.

Land Information New Zealand. 2008. *Standard for New Zealand Geodetic Datum 2000 Projections: Version 2 - LINZS25002*. Wellington: Land Information New Zealand. URL http://www.linz.govt.nz/regulatory/25002.

Land Information New Zealand. 2010. *Rating Valuations Rules 2008: version date 1 October 2010 - LINZS30300*. Wellington: Land Information New Zealand. URL http://www.linz.govt.nz/regulatory/30300.

Lazrak, F., Nijkamp, P and Rietveld, P. 2014. The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. Journal of Geographic Systems, 16(1), pp. 89-114.

Lees, K. 2014. “Big city life? Challenges and trade-offs for Auckland city.” Wellington: New Zealand Institute of Economic Research, working paper 2014/02.

Lees, K. 2015. “Moving on up: Relaxing land use restrictions can lift Auckland city.” Wellington: New Zealand Institute of Economic Research, report to Auckland Council.

McGreal, S and de La Paz, P. 2013. Implicit house prices: Variation over time and space in Spain. Urban Studies, 50(10), pp. 2024-2043.

McMillen, D.P and McDonald, J. 2002. “Land Value in a Newly Zoned City.” *The Review of Economics and Statistics, Vol.* 84, No. 1, pp. 62-72.

MRCagney. 2013. The Economic Impacts of Parking Requirements in Auckland, Auckland: MRCagney on behalf of Auckland Council.

MRCagney. 2014. “The economic impacts of minimum apartment and balcony rules.” Auckland, MRCagney and Auckland Council.

New Zealand Productivity Commission. 2012. Housing affordability inquiry: Final report. Wellington: New Zealand Productivity Commission.

Pebesma, E.J and Bivand, R.S. 2005. “Classes and methods for spatial data in R.” *R News* 5 (2). URL http://cran.r-project.org/doc/Rnews/.

Rawlinsons. 2014. *Rawlinsons New Zealand Construction Cost Handbook 2013/14*, 28th Edition. Auckland: Rawlinsons Publications.

Rehm, M. 2009. “Judging and house by its cover: Leaky building stigma and house prices in New Zealand.” *International Journal of Housing Markets and Analysis*, 2(1).

Rehm, M and Filippova, O. 2008. “The impact of geographically defined school zones on house prices in New Zealand.” *International Journal of Housing Markets and Analysis* 1(4).

Rehm, M., Filippova, O and Stone, J. 2006. “The influence of vintage on house value.” *Pacific Rim Property Research Journal* 12(3).

Roback, J. 1982. “Wages, Rents and the Quality of Life.” *Journal of Political Economy*, 90(6).

Rohani, M. 2012. “Impact of Hauraki Gulf amenity on the land price of neighbourhood properties. An empirical Hedonic Pricing Method case study, North Shore, Auckland.” Auckland: Auckland Council Working Paper 2012/001.

Rosen, S. 1979. “Wage-Based Indices of Urban Quality of Life”, in P. Miezkowski and M. Strazheim (eds.), *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press.

Samarasinghe, O and Sharpe, B. 2010. “Flood prone risk and amenity values: a spatial hedonic analysis.” *The Australian Journal of Agricultural and Resource Economics*, 54(1), pp. 457-475.

Sheppard, S. 1999. “Hedonic Analysis of Housing Markets.” In *Handbook of Regional and Urban Economics*, eds E.S. Mills and P. Cheshire.

Statistics New Zealand. 2014. *2013 Census meshblock dataset*. Wellington: Statistics New Zealand. URL http://www.stats.govt.nz/Census/2013-census/data-tables/meshblock-dataset.aspx.

Statistics New Zealand. 2015. *Labour Market Statistics*. Wellington: Statistics New Zealand. URL http://www.stats.govt.nz/browse_for_stats/income-and-work/employment_and_unemployment/labour-market-statistics-information-releases.aspx.

Torshizian, E. 2014. “SPWT: A Stata Command for Creating Spatial Weight Matrices”, Auckland: University of Auckland Department of Economics Working Paper.

Torshizian, E and Grimes, A. 2014a. “Residential Satisfaction, Crowding and Density: Evidence over Different Geographic Scales in Auckland.” Auckland: New Zealand Association of Economists 2014 Conference. URL: http://www.nzae.org.nz/wp-content/uploads/2015/01/Residential_Satisfaction_Crowding_and_Density_ET.pdf

Torshizian, E and Grimes, A. 2014b. “Negative Envy or Positive Amenity Effects? A Neighbourhood Study of Aucklanders’ Residential Satisfaction”. London: Regional Studies Association Winter Conference 2014. URL http://www.regionalstudies.org/uploads/conferences/presentations/Winter_2014_Presentations_/Eilya_Torshizian_and_Arthur_Grimes.pdf

Wieckowski, A. 2010. “Back to the City”. *Harvard Business Review* May 2010 issue.

Zeileis, A and Hothorn, T. 2002. “Diagnostic Checking in Regression Relationships.” *R News* 2(3), 7-10. URL http://CRAN.R-project.org/doc/Rnews/.

Zheng, G. (2013). *The effect of Auckland's metropolitan urban limit on land prices*. Wellington, New Zealand Productivity Commission.

# 7.0 Appendix: Additional data and analysis

Table 4: Identification of preferred OLS regression model

| Auckland residential property sales 2011-2014, OLS models | | | | |
|---|---|---|---|---|
| | *Dependent variable:* log(sale_price_net) | | | |
| | Preferred OLS model | Alternative model 1 (exhibits multicollinearity) | Alternative model 2 (excl. building quality) | Alternative model 3 (excl. neighbourhood variables) |
| log(LAND) | $0.121^{***}$ | $0.122^{***}$ | $0.109^{***}$ | $0.133^{***}$ |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| log(FLOORSPACE) | $0.611^{***}$ | $0.610^{***}$ | $0.639^{***}$ | $0.663^{***}$ |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| log(DCBD) | $-0.283^{***}$ | $-0.284^{***}$ | $-0.273^{***}$ | $-0.310^{***}$ |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| log(DCOAST) | $-0.011^{***}$ | $-0.012^{***}$ | $-0.011^{***}$ | $-0.009^{***}$ |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| PRE1940 | $0.118^{***}$ | $0.119^{***}$ | $0.103^{***}$ | $0.169^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| CARPARKS | $0.001^{***}$ | $0.001^{***}$ | $0.001^{***}$ | $0.002^{***}$ |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| VIEW:OTHER (1) | $0.006^{**}$ | $0.006^{**}$ | $0.008^{***}$ | $0.018^{***}$ |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| VIEW:WATER (1) | $0.172^{***}$ | $0.172^{***}$ | $0.170^{***}$ | $0.210^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| COND_WALL:FAIR (2) | 0.001 | -0.01 | | -0.004 |
| | (0.009) | (0.01) | | (0.009) |
| COND_WALL:GOOD (2) | $0.072^{***}$ | $0.048^{***}$ | | $0.089^{***}$ |
| | (0.003) | (0.005) | | (0.003) |
| COND_WALL:POOR (2) | $-0.141^{***}$ | $-0.093^{***}$ | | $-0.147^{***}$ |
| | (0.015) | (0.021) | | (0.016) |
| COND_WALL:MIXED (2) | 0.006 | 0.019 | | 0.001 |
| | (0.021) | (0.028) | | (0.022) |
| COND_ROOF:FAIR (2) | | $0.027^{***}$ | | |
| | | (0.01) | | |
| COND_ROOF:GOOD (2) | | $0.028^{***}$ | | |
| | | (0.005) | | |
| COND_ROOF:POOR (2) | | $-0.087^{***}$ | | |
| | | (0.023) | | |
| COND_ROOF:MIXED (2) | | -0.024 | | |
| | | (0.03) | | |
| MBHERITAGE | $0.004^{***}$ | $0.004^{***}$ | $0.005^{***}$ | |
| | (0.0002) | (0.0002) | (0.0002) | |
| log(HHINCOME) | $0.201^{***}$ | $0.201^{***}$ | $0.209^{***}$ | |
| | (0.004) | (0.004) | (0.004) | |
| log(DENSITY) | $-0.038^{***}$ | $-0.038^{***}$ | $-0.039^{***}$ | |
| | (0.001) | (0.001) | (0.001) | |
| SALEYEAR:2012 (3) | $0.052^{***}$ | $0.052^{***}$ | $0.052^{***}$ | $0.042^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| SALEYEAR:2013 (3) | $0.186^{***}$ | $0.186^{***}$ | $0.186^{***}$ | $0.175^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| SALEYEAR:2014 (3) | $0.267^{***}$ | $0.267^{***}$ | $0.266^{***}$ | $0.257^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Constant | $11.008^{***}$ | $11.019^{***}$ | $10.700^{***}$ | $13.174^{***}$ |
| | (0.053) | (0.053) | (0.051) | (0.031) |
| Observations | 72,855 | 72,855 | 72,855 | 72,855 |
| Adjusted $R^2$ | 0.640 | 0.640 | 0.636 | 0.615 |
| F Statistic | $7,197.600^{***}$ (df = 18; 72836) | $5,895.547^{***}$ (df = 22; 72832) | $9,096.188^{***}$ (df = 14; 72840) | $7,759.572^{***}$ (df = 15; 72839) |
| **Akaike Information Criterion score** | **37478.48** | **37431.65** | **38272.99** | **42373.78** |

Notes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

(1) VIEW is a categorical (dummy) variable. The base level is "no view".
(2) COND_WALL and COND_ROOF are a categorical (dummy) variables. The base level is "average" condition.
(3) SALEYEAR is a categorical (dummy) variable. The base level is 2011.

Table 5: Comparison of OLS models on full dataset and randomly selected subset of 10,000 sales

| **Auckland residential property sales 2011-2014, OLS models** | | |
|---|---|---|
| | *Dependent variable:* | |
| | log(sale_price_net) | |
| | Full dataset | Subset |
| log(LAND) | 0.121*** | 0.120*** |
| | (0.002) | (0.007) |
| log(FLOORSPACE) | 0.611*** | 0.609*** |
| | (0.003) | (0.009) |
| log(DCBD) | -0.283*** | -0.284*** |
| | (0.002) | (0.006) |
| log(DCOAST) | -0.011*** | -0.014*** |
| | (0.001) | (0.003) |
| PRE1940 | 0.118*** | 0.121*** |
| | (0.004) | (0.011) |
| CARPARKS | 0.001*** | 0.001 |
| | (0.0004) | (0.001) |
| VIEW:OTHER (1) | 0.006** | 0.006 |
| | (0.003) | (0.008) |
| VIEW:WATER (1) | 0.172*** | 0.164*** |
| | (0.004) | (0.011) |
| COND_WALL:FAIR (2) | 0.001 | -0.022 |
| | (0.009) | (0.022) |
| COND_WALL:GOOD (2) | 0.072*** | 0.066*** |
| | (0.003) | (0.008) |
| COND_WALL:POOR (2) | -0.141*** | -0.168*** |
| | (0.015) | (0.041) |
| COND_WALL:MIXED (2) | 0.006 | -0.006 |
| | (0.021) | (0.053) |
| MBHERITAGE | 0.004*** | 0.004*** |
| | (0.0002) | (0.0005) |
| log(HHINCOME) | 0.201*** | 0.194*** |
| | (0.004) | (0.01) |
| log(DENSITY) | -0.038*** | -0.034*** |
| | (0.001) | (0.004) |
| SALEYEAR:2012 (3) | 0.052*** | 0.068*** |
| | (0.004) | (0.01) |
| SALEYEAR:2013 (3) | 0.186*** | 0.197*** |
| | (0.004) | (0.01) |
| SALEYEAR:2014 (3) | 0.267*** | 0.300*** |
| | (0.004) | (0.011) |
| Constant | 11.008*** | 11.093*** |
| | (0.053) | (0.144) |
| Observations | 72,855 | 10,000 |
| Adjusted $R^2$ | 0.640 | 0.625 |
| F Statistic | 7,197.600*** (df = 18; 72836) | 927.097*** (df = 18; 9981) |
| *Notes:* | **p<0.1; **p<0.05; ***p<0.01* | |

(1) VIEW is a categorical (dummy) variable. The base level is "no view".
(2) COND_WALL is a categorical (dummy) variable. The base level is "average" condition.
(3) SALEYEAR is a categorical (dummy) variable. The base level is 2011.