Commuting and Residential Decisions in the Greater Wellington Region

Toby Daglish, Mairéad de Róiste, Yiğit Sağlam, and Richard Law

March 3, 2015

Abstract

This paper studies residential, commuting and car ownership decisions in the Greater Wellington Region of New Zealand. We establish an estimation methodology that is robust to endogeneity between house prices and residential decisions. The paper also makes extensive use of Geographic Information Systems calculations, allowing us to evaluate the impact of schools, greenspaces and sunlight on decisions. The paper finds that commuting decisions are highly affected by demographic variables, that amenities are important in determining neighbourhood preferences, and that school quality, contrary to popular belief, has relatively little effect on decisions.

Households typically face a three-way decision on moving to a new region: where to live, how to commute to work, and how many cars to own. The drivers of this decision are critically important for planners and policy-makers, who must decide how to provide roading infrastructure, public transport services, and pedestrian routes, along with how to incentivise commuters to behave so as to minimise congestion problems. These decisions also have an important influence on the development of cities, since commuting preferences shape the demand for convenient downtown apartments versus spacious suburban sprawl.

Our paper's contribution to the commuting literature is three-fold. First, we develop a careful methodology for the treatment of endogeneity of house prices. Second, we examine a relatively unexplored dataset, which is potentially more interesting than single-city transport surveys. Third, we make extensive use of Geographic Information Systems (GIS) computations, allowing us to include variables that would not otherwise be available in transportation studies.

Endogeneity is a problem that pervades models of residential choice: households prefer houses with desirable characteristics, but these same characteristics drive up house prices, which are undesirable. This problem is discussed in further detail by Blundell and Powell (2004), Guevara and Ben-Akiva (2006), and Petrin and Train (2010). We develop an iterative estimation, using estimates of demand for houses based upon our logistic model for residential/commuting choice to explain house prices exogenously, and then feed house price residuals into our logistic model to ensure that we are capturing exogenous movements in house prices. Empirically, we find that our estimates produce economically sensible price/time sensitivities.

The data used in this study covers the Greater Wellington Region, using the New Zealand Ministry of Transport's Household Travel Survey (HTS). This annual survey selects a panel of households around the country, who keep a diary of their movements for a period of two days. Using this information, we are able to observe where households live, and how they commute to work. Demographic information is also provided, along with information on car ownership. In contrast to previous studies (such as Salon (2009) and Bhat and Guo (2007)), our data set has a time series component to it, allowing a richer analysis of time varying variables on commuting decisions. In particular, evolving house prices may provide push and pull factors in explaining the time series behaviour of residential decisions. In addition, since our survey covers an entire region, as opposed to a single city, our data has considerable variety to it: as well as commuters travelling to the downtown core, we also see commuters in more provincial areas commuting to local business districts.

Our work makes extensive use of explanatory variables generated using Geographic Information Systems (GIS) software. Transport is not aspatial. To understand the factors influencing commuters, we need to understand and model spatial factors that affect their use of space. With the increase in data availability and processing power, GIS computations facilitate accurate modelling of transport networks. These spatial models provide realistic distance and cost attributes, which affect commute mode, car ownership and residential decisions. These also allow us to consider closest schools, availability of sunlight, and proximity to green spaces as drivers of residential decisions.

The layout of the remainder of this paper is as follows: Section 1 describes our methodology, Section 2 discusses our data, and Section 3 outlines our results. Finally, Section 4 concludes.

1 Methodology

1.1 Conditional Logit Estimation

Our model assumes a standard Conditional Logit model for individual preferences over modal/residential choices. Individual i's utility from choice j is given by:

$$u_{i,j} = \sum_{k=1}^{K} \beta_k x_{i,j,k} + \epsilon_{i,j}$$

where β_k $(k = 1, \ldots K)$ are assumed constant, $x_{i,j,k}$ is a covariate used to explain person i's preferences, and $\epsilon_{i,j}$ is assumed to be logistically distributed. Under these assumptions, the probability that person i chooses choice j is given by:

$$\text{Prob}_i(j) = \frac{e^{\sum_k \beta_k x_{i,j,k}}}{\sum_l e^{\sum_{k'} \beta_{k'} x_{i,l,k'}}}.$$

From this, we can calculate the likelihood of each of the individuals making the decision observed in the data. We then choose β_k $(k = 1, \ldots, K)$ to maximise this likelihood:

$$\mathcal{L} = \prod_i \text{Prob}_i(j(i)),$$

where $j(i)$ is the choice individual i made, as observed in the data. In considering the choice of where to live, we allow an individual to choose from a selection of randomly selected meshblocks, chosen in a stratified fashion, so that one meshblock is available in each area unit of the study (see Section 2).

1.2 Endogeneity

We note that problems with estimation of Conditional Logit models frequently occur in residential choice models due to endogeneity of house prices. Since house prices are generally higher in desirable areas, omitted variables can lead to under-estimation of the sensitivity of individual utility to house prices, potentially leading to (mis-estimated) dis-utility from cheaper housing. To control for this, we use a collection of instruments (see section 2.4) for house prices. In this set of instruments, we include a measure of demand derived from individual residential choice probabilities:

$$\text{Demand}_k = \frac{\sum_i \sum_{j \in \mathcal{J}(k)} \text{Prob}_i(j)}{\#(\mathcal{J}(k))}, \tag{1}$$

where we define the set $\mathcal{J}(k)$ as being those choices where the individual would live in location k, and $\#(x)$ being the number of elements in set x.[1]

Since this demand variable is itself a function of our estimated β coefficients, our estimation requires us to solve the simultaneous system:

$$\max_{\beta} \prod_i \text{Prob}_i(j(i)|v) \quad P = Z\gamma + v \tag{2}$$

where Z is a collection of instruments (including Demand), v is the collection of residuals

[1] We scale the probabilities to account for differences in numbers of meshblock per area unit that may result in some meshblocks being over-sampled in our stratified sampling.

from regressing house prices (P) on Z and

$$\mathrm{Prob}_i(j(i)|v) = \frac{e^{\delta v(j)+\sum_k \beta_k x_{i,j,k}}}{\sum_l e^{\delta v(l)+\sum_{k'} \beta_{k'} x_{i,l,k'}}}, \tag{3}$$

where $v(j)$ is the house price residual associated with the residential location for choice j.[2]

The two equations (2) can be used to generate a mapping on the space of model parameters β: given β, we can generate probabilities through (3), which in turn allow us to generate demand through (1). This in turn allows us to estimate a new vector for β by maximum likelihood using (3). The algorithm has converged when β maps to itself, i.e.

$$\beta^* = \arg\max_{\beta} \prod_i \frac{e^{\delta v(k|\beta^*)+\sum_k \beta_k x_{i,j,k}}}{\sum_l e^{\delta v(k'|\beta^*)+\sum_{k'} \beta_{k'} x_{i,l,k'}}} \tag{4}$$

where we emphasise the dependence of v on β by denoting the relevant $v(k|\beta)$. Realising that this is a nonlinear equation in the vector β^*, we use Broyden's method to solve (4), where each evaluation of (4) requires an evaluation of (2). This iteration normally converges after a small number of iterations, and is quite computationally viable for large problems.

2 Data

2.1 The commuters

The New Zealand Household Travel Survey (HTS) is conducted by the Ministry of Transport (MoT) and records trip choices of 2,200 (for 2004–2008) or 4,600 households (for 2009-

[2]Note that our house price estimation is performed as a panel regression over time and meshblock. Hence $v(k)$ is a function of the particular year in which individual i makes his/her decision.

present) across New Zealand. Each person within each household is asked to record trips taken across two consecutive travel days and are then interviewed about their travel behaviour and relevant socio-economic and demographic characteristics.[3]

Our work explores the trip choices made by HTS participants within the Greater Wellington Region (GWR). The region's population is just over 470,000 in the 2013 census. The GWR comprises 9 local authority areas and contains four of New Zealand's 13 cities, ranging in population from 52,000 to 191,000. The region is located on the South West of the North Island, and is constrained by the Cook Strait, a natural harbour, and hilly terrain. The nation's capital, Wellington City, is the main urban area. The region is served by suburban and intercity rail, a harbour commuter ferry, cable car, and buses. The main trunk roads, State Highways 1 and 2, are confined by the terrain and coastline.

The HTS participants included in this studyare commuters who live within GWR and work in GWR, at least 10 kilometres from the regions boundary. We exclude work locations near the boundary, since, the residential choice set for these commuters will include many locations outside the study area. Participants are further filtered by those with a single work location and erroneous records are removed. While employment type is not identified in the HTS, in this way participants with changeable work locations, such as tradespeople who are unlikely to optimise their residential location according to their work locations, are excluded from our model. This data filtering process identifies 1115 participants in 769 households. Figure 1 shows the preponderance of drivers in the Wellington region.

MOT provides our project with the X and Y coordinates of each participants home and

[3]Detailed survey information is available at:
http://www.transport.govt.nz/research/travelsurvey/detailedtravelsurveyinformation/.

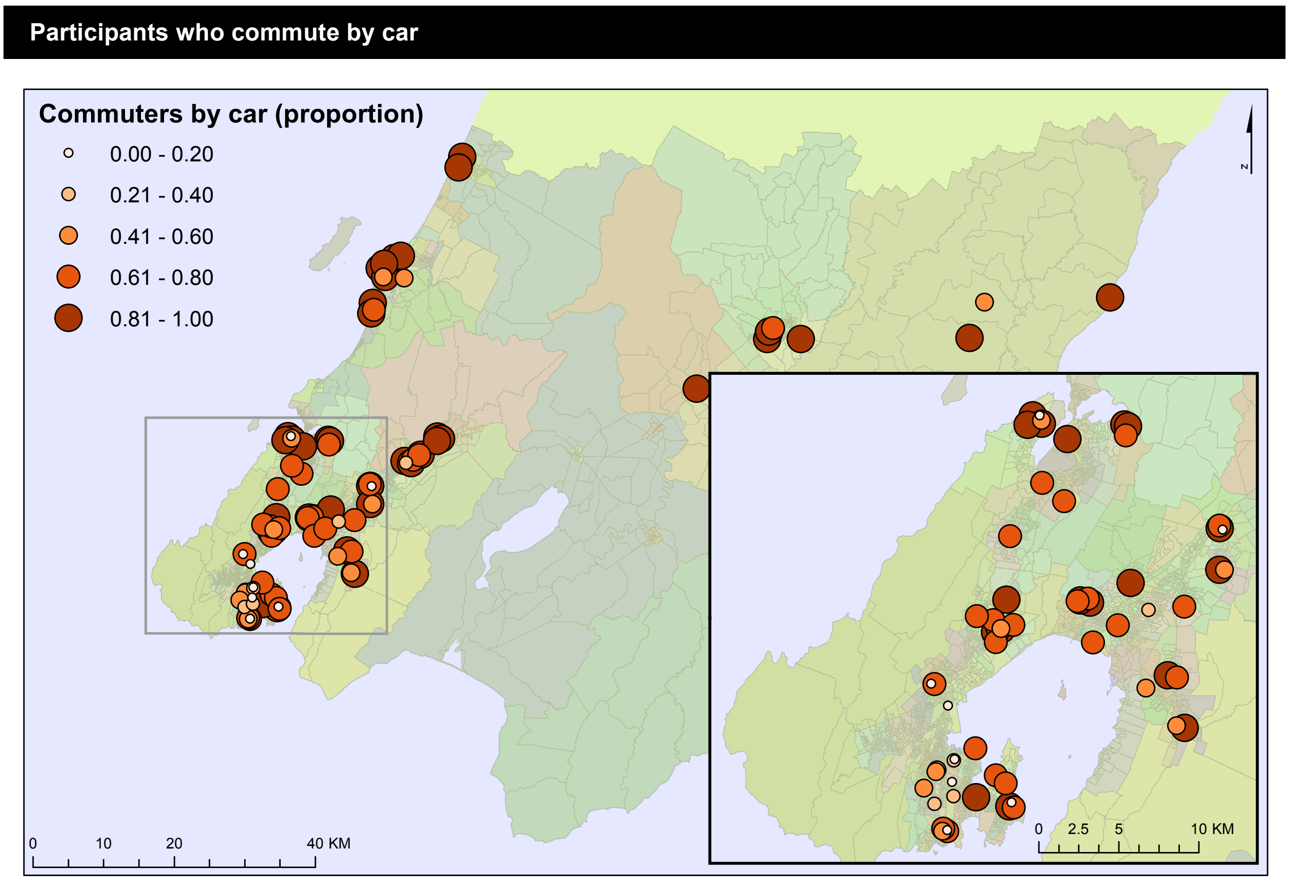

Figure 1: Proportion of drivers by meshblock in Greater Wellington Region. Circles illustrate meshblocks sampled by the HTS. Size of circle illustrates proportion of commuters in meshblock who commute by driving.

work locations for our study area. The locations are matched to meshblock areas (the smallest census geographic unit for New Zealand with a population range of normally between 60–110 residents). Each home location is then anonymised to a point in the relevant meshblock. The meshblock points are created by weighting all possible address locations in the meshblock to find the most central address point, and then moving this point to the nearest location on the road network. We use these central addresses for the participants' home locations, as well as other possible alternatives to enable appropriate comparison.

Meshblocks are amalgamated into census Area Units (AU) wholly contained within the region. Each of the 194 terrestrial AUs in the GWR ranges in size from 150km^2 to 1,664,988km^2 and contain between 1 and 71 meshblocks. AUs normally contain 3-5,000 residents, aggregate to urban or rural areas and larger administrative units and are roughly analogous to suburbs. To generate our alternative residential locations, each household is randomly allocated a meshblock within every AU they did not reside in. This is then added to their actual residential meshblock to create a set of 194 possible residential choices. The AUs vary considerably over the region and provide a varying choice set for each household from large rural properties in tight knit communities to urban apartments in the city centre.

2.2 The network

This section describes first, the creation of our multimodal network (driving, walking, cycling and taking public transport) and the spatial factors affecting the cost and distance calculations for each of these networks. Second, we cover the spatial factors of different residential locations which feed into residential decisions; green space, house prices, sea views, access to

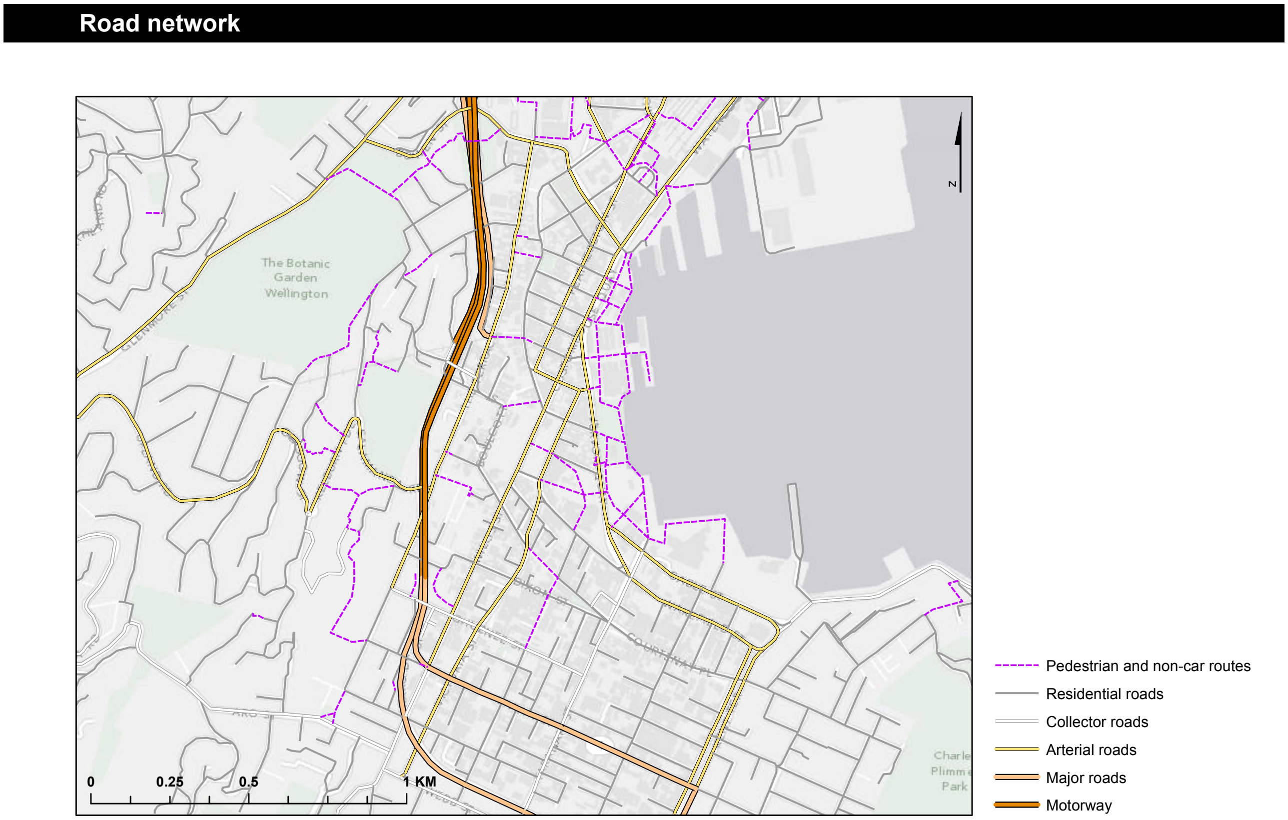

Figure 2: Downtown Wellington road network, illustrating the preponderance of pedestrian shortcuts incorporated in our modelling. Active Transport travel times are substantially affected by the ability to move by means other than the road network.

amenities, and schools.

Our model assesses travel over four possible networks: walking, cycling, driving, and public transport. Modal choice is further increased by the possibility of driving or walking to public transport.

The transport network is adapted from a purchased road dataset, based on NZ Open GPS data. The network is clipped to the study area, checked for errors and omissions and updated. Pedestrian routes are added. Although the multiple sources consulted to error check and update the network means that exact dating is problematic, the network primarily dates to 2011. Figure 2 shows the portion surrounding the downtown area, illustrating pedestrian routes. Driving travel time over the network is calculated using speed restrictions while pedestrian and cycle travel times are calculated based on distance and hill slope (a particularly important factor in Wellington's hilly terrain). Motorists are assumed to follow the appropriate speed limit. Each route within the network is restricted according to one way constraints and permitted uses, e.g. pedestrians are not permitted to walk on a motorway and cars are unable to take pedestrian shortcuts. The public transport network requires further consideration as public transport commuters are only able to enter and exit the public transport network at defined points, e.g. bus stops or train stations. The travel times for public transport can vary significantly by time of day. Consequently, non-premium, peak travel times (7–9am) are taken from publicly available timetables (the Google Transit Feed Specification; GTFS). Waiting times for public transport are modelled as a constant. Travel times are calculated using ArcGIS Network Analyst's shortest path based on Dijkstra's algorithm. Travel costs are also captured for this route.

2.3 Housing

Property values are influenced by a number of factors with distinct spatial patterns.

2.3.1 House prices and vintage

House price data is supplied by QV for each meshblock for each study year. We use three numbers here for each meshblock. The median Crown Valuation (CV) provides a measure of the accounting value of a typical home in the area. The median sales price and median

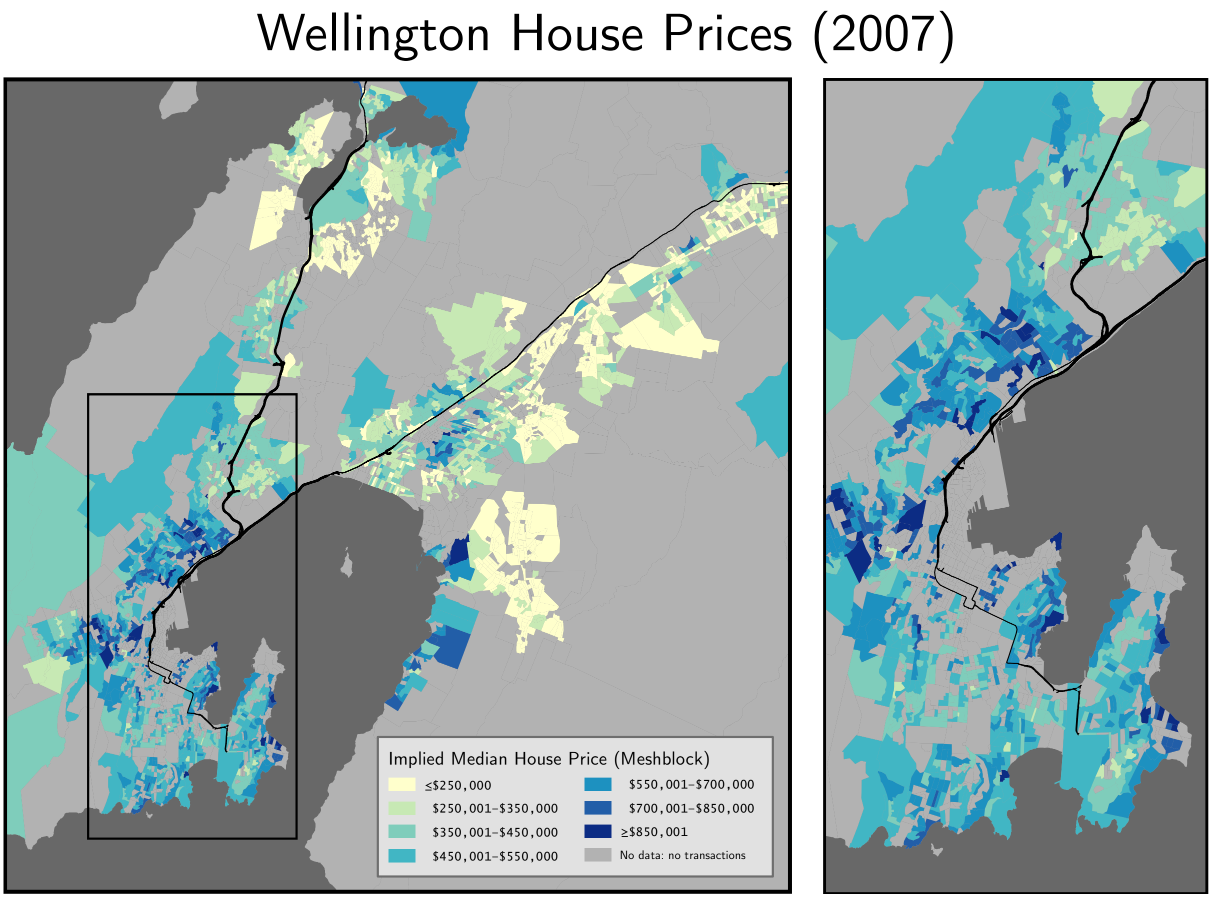

Figure 3: House prices in Wellington City, Hutt Valley, and Porirua, 2007. Note the strong correlation between house prices and proximity to the downtown core.

CV of house sold allows us to create an inflator to convert CV to market valuation. We apply this inflator to the median CV of the entire population of houses in the Meshblock in order to create our measure of house price in the area. This avoids fluctuations in house prices caused by different subsets of houses changing hands each year; given the small size of Meshblocks, often only 1-5 houses will sell in a given year. Figure 3 illustrates house prices in the Wellington and Lower Hutt regions for the year 2011.

2.3.2 Industrial and Commercial

Commercial and Industrial land uses were determined from the six constituent land use plans in our study area. While the definitions of land use type varies across the different local authorities, the data was homogenised as much as possible. The area in each meshblock and area unit for both industrial and commercial land use was calculated. In our study, this value determines whether the immediate neighbourhood (meshblock) or wider suburb is dominated is dominated by either land use type.

2.3.3 Green space

Green space proximity has been linked to property prices by Comber, Brunsdon, and Green (2008), Conway, Li, Wolch, Kahle, and Jerrett (2010), Higgs, Fry, and Langford (2012), and Kaufman and Cloutier (2006), suggesting that green spaces could well be an important consideration by households choosing a residential location.

The quality and location of green space within the study area are classified using a Normalised Difference Vegetation Index (NDVI) in ENVI software on Landsat 5 imagery for the 8 of November 2005. Two Landsat multispectral, 30m resolution images are used with 10% cloud coverage (with cloud coverage primarily over the ocean areas). Vegetation in the study area is divided into four categories; none, sparse, and dense vegetation. The percentage of land in each meshblock and AU covered by each of the four vegetation classes is then calculated. Figure 4 demonstrates greenspace evaluation in the downtown area.

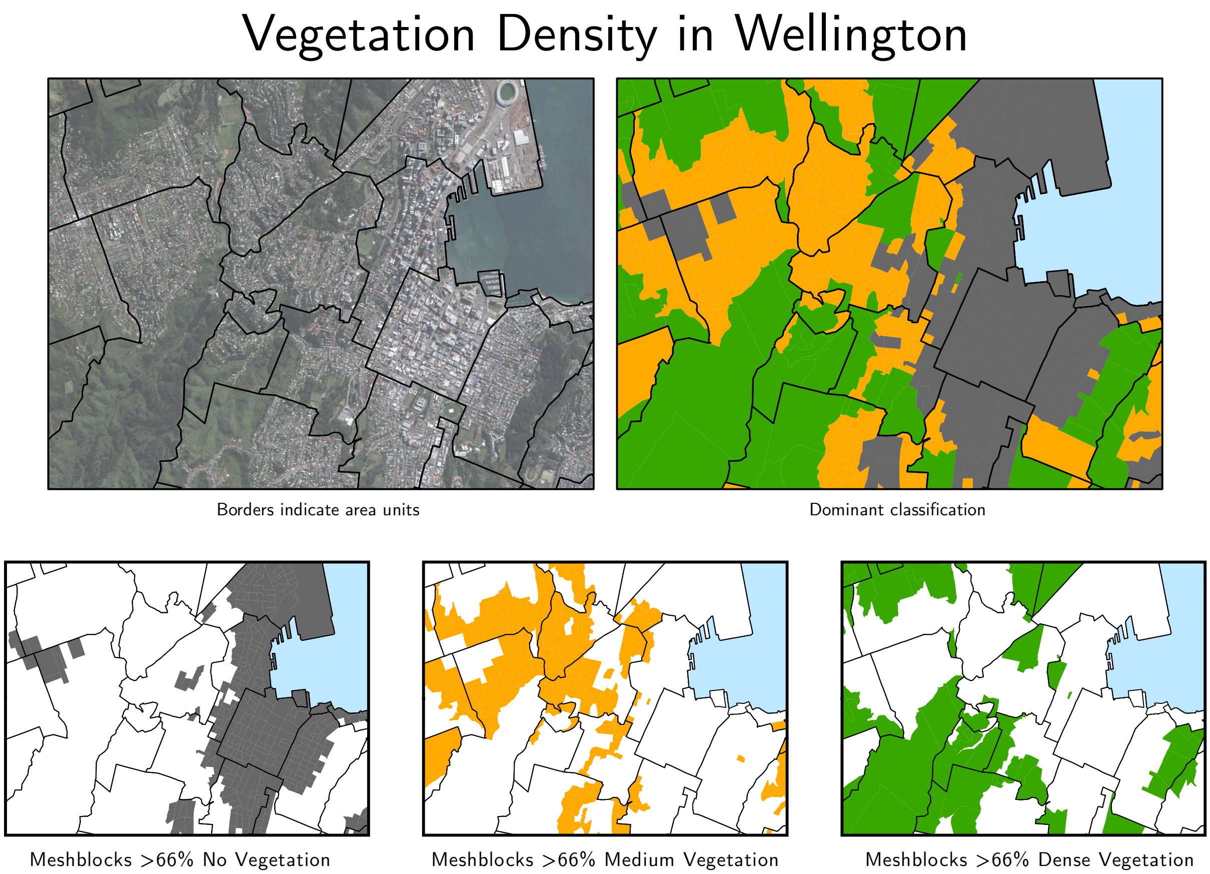

Figure 4: Greenspace classification in the downtown area. First graph shows satellite image of downtown. Second graph shows classification into vegetation type. Lower graphs show meshblocks classified by predominate type of vegetation.

2.3.4 Sunlight

Sunlight is less often considered in residential decision modelling but has been found to have a significant effect on house prices (Helbich, Jochem, Mücke, and Höfle (2013)). Residential land is combined from the district plans of the constituent local authorities. In the case of Porirua City Council, this information is not available, and is created through the manipulation of other zoning information and a 40m buffer of all residential address points. As this study is conducted in the Southern Hemisphere, North, North East and North West facing slopes (0 to 67.5 degrees and 292.5 to 360 degrees) are used as a proxy for sunlight. Slopes with this aspect are classified, based on a 15m DEM and the proportion of residential area in each meshblock meeting this criteria is recorded.

2.3.5 Schools

A number of schools within the study area operate a zoning system to limit student enrolment. The school zones are compiled for 2010. Locations of all 256 primary, intermediate and secondary schools within the study area are originally acquired from koordinates.com, and missing observations are added.[4] These school features are then moved to the closest road, and the driving time from each meshblock centroid to each school along the road network is calculated. For each potential residential location, we calculate the closest zoned primary school, boys' secondary school, girls' secondary school, and co-educated secondary school. For secondary schools, we then generate a boy's secondary school measure as being based upon either the performance of the closest boys' school or the closest co-educated

[4] https://koordinates.com/layer/243-nz-schools/

school, whichever has the shortest travel time. An analogous number is calculated for girl's secondary school performance. Finally, we assume that household's may have a preference for one or the other based on the gender mix of their children. Assuming that households self-select according to this, we then measure secondary school performance as being the higher of the boy's performance measure and the girl's performance measure. Note that GWR has relatively few single sex secondary schools, so in most cases, the closest school is a co-educated school, and therefore boy's and girl's performance is identical.

2.4 Instruments

In explaining house prices, we have a large number of degrees of freedom (we observe each meshblock in Wellington most years; only missing it when no houses change hands that year), and therefore use a large set of instruments to explain house prices. We use: year dummy variables, closest primary school decile, closest girls', boys' and co-educated secondary school University Entrance pass rates, the average of neighbouring meshblock house prices, our demand measure, and the number of houses in the meshblock. This last variable gives a measure of supply in the meshblock. In general, meshblocks are designed so that there are approximately fifty houses in each. However, due to population growth, a "pregnant" meshblock may develop in some areas, where new houses have been built, but the meshblock has not yet been split in two. Conversely, a shrinking meshblock or recently split meshblock may be smaller than average.

3 Results

3.1 Summary Statistics

Table 1 contains similar information for the HTS participants. Several points are worthy of note. First, the preponderance (64%) of household heads are men. The ethnic mix is roughly similar to New Zealand's ethnic mix, and is therefore unremarkable. Family sizes vary considerably, as do the number of working adults. Most household heads (96%) have drivers licences. The average commuter lives 13 minutes from their work by car, 52 minutes by bicycle, or 4.5 hours on foot. Public transport is slightly more complex. If a commuter drove to a park-and-ride facility (generally at a train station) their commute would be 44 minutes, but if they walked to a bus or train and then rode public transport, their commute time would be 52 minutes (similar to cycling). The average neighbourhood house price is around $380 000, with considerable dispersion. Here we should emphasise that house prices reported are at time of survey participation, so some "low" prices are associated with earlier times (see Table 2.

School performance is mixed. University entrance exam scores vary considerably across schools in the region, with the two girls' schools present scoring well. Primary school deciles cover the entire spectrum, but most survey respondants live close to schools that are slightly above average by national standard (mean decile of 6.89).

Respondents tend to live in sparsely vegetated areas, with very little commercial or industrial land near to them, although there si considerable variation here. There is also considerable variation in terms of sunlight in their homes, with the average respondent living in a meshblock that is 40% north facing.

	Mean	SD	Min	Max	Nnan	Nobs
D_female	0.36	0.48	0	1	0	624
age	43.64	11.95	16	76	0	624
D_peincome	10.16	3.33	1	17	0	624
D_maori	0.06	0.25	0	1	0	624
D_pasifika	0.03	0.17	0	1	0	624
D_asia	0.08	0.27	0	1	0	624
N_kids10	0.42	0.78	0	5	0	624
N_kids18	0.28	0.65	0	5	0	624
N_adultw	1.36	0.68	0	4	0	624
D_peclicn	0.96	0.2	0	1	2	624
Min_Walk	274.08	274.43	0.17	1963.98	0	624
Min_Cycle	51.65	51.65	0.04	421.01	0	624
Min_Car	13.33	11.8	0.01	105.19	0	624
Min_PTCar	44.35	29.56	2.83	582.86	0	624
Min_PTWalk	51.87	56.04	0.17	857	0	624
WorkWLGDT	0.44	0.5	0	1	0	624
WorkLHDT	0.02	0.15	0	1	0	624
WorkPorDT	0.03	0.17	0	1	0	624
Price_Act	380195.23	165078.96	86194.61	1093951.99	64	624
ID_Act	6.89	2.77	1	10	0	624
UE_Coed_Act	0.34	0.14	0	0.65	0	624
UE_Boys_Act	0.32	0.12	0.13	0.74	0	624
UE_Girls_Act	0.71	0.13	0.37	0.87	0	624
DT_Prim_Act	1.47	1.05	0.09	8.51	0	624
DT_Coed_Act	4.02	3.07	0.64	26.85	0	624
DT_Boys_Act	26.16	21.69	0.03	112.06	0	624
DT_Girls_Act	21.19	19.86	1.57	105.01	0	624
PercMBNoVeg_Act	27.18	22.38	0	100	0	624
PercMBSparseVeg_Act	53.52	20.72	0	94.44	0	624
PercMBDenseVeg_Act	19.3	21.82	0	94.94	0	624
PercAUNoVeg_Act	24.59	17.99	0.51	84.92	0	624
PercAUSparseVeg_Act	42.4	14.32	6.94	71.76	0	624
PercAUDenseVeg_Act	33	20.69	0.78	91.66	0	624
MBMeanBed_Act	3.07	0.38	2.18	4.4	14	624
MB_Prop_Comm_Act	0.02	0.07	0	0.69	0	624
AU_Prop_Comm_Act	0.02	0.04	0	0.56	0	624
MB_Prop_Ind_Act	0.06	0.16	0	0.75	0	624
AU_Prop_Ind_Act	0.01	0.02	0	0.15	0	624
PercNorth_Act	40.71	30.03	0	100	0	624

Table 1: Summary statistics for Household heads. GIS variables report the observed ammenities of participants actual residential locations. Drive times are listed for schools, as are the optimal travel times for each choice of travel: AT, Driving, and PT either using a car to drive to a park and ride location, or walking.

	Mean	SD	Min	Max	Nnan	Nobs
PrY03	243449.52	136920.96	43828.13	2266313.29	1602	4682
PrY04	262663.31	140641.60	46952.04	1304527.61	1633	4682
PrY05	283746.22	150732.49	46504.31	1744913.79	1618	4682
PrY06	333845.44	172210.11	59967.86	2884090.91	1650	4682
PrY07	390096.49	200866.62	65303.7	2505952.38	1718	4682
PrY08	421672.34	196089.73	61966.55	3760973.01	2217	4682
PrY09	429886.84	200622.39	78069.54	2914229.63	2044	4682
PercMBNoVeg	32.53	32.38	0	100	0	4682
PercMBSparseVeg	45.61	28.66	0	100	0	4682
PercMBDenseVeg	21.86	29.19	0	100	0	4682
PercAUNoVeg	25.59	24.19	0	100	0	4682
PercAUSparseVeg	36.36	16.12	0	71.76	0	4682
PercAUDenseVeg	38.04	24.5	0	91.66	0	4682
MBMeanBed	3	0.45	1	4.5	1208	4682
MB_Prop_Comm	0.04	0.16	0	1	3	4682
AU_Prop_Comm	0.04	0.12	0	0.69	3	4682
MB_Prop_Ind	0.08	0.18	0	0.75	3	4682
AU_Prop_Ind	0.02	0.07	0	0.6	3	4682
PercNorth	33.39	32.71	0	100	3	4682
ID03	6.2	3.1	1	10	181	4682
ID04	6.2	3.06	1	10	181	4682
ID05	6.28	3.13	1	10	181	4682
ID06	6.27	3.13	1	10	181	4682
ID07	6.27	3.13	1	10	181	4682
ID08	6.26	3.13	1	10	181	4682
ID09	6.26	3.12	1	10	181	4682
UE_Coed_03	18.3	12.18	0	44	181	4682
UE_Coed_04	27.66	12.21	6	45	181	4682
UE_Coed_05	28.70	12.77	8	47	181	4682
UE_Coed_06	27.14	12.14	8	48	181	4682
UE_Coed_07	32.38	12.82	0	53	181	4682
UE_Coed_08	36.39	12.87	13	61	181	4682
UE_Coed_09	34.95	13.77	10	65	181	4682
UE_Boys_03	20.91	18.25	13	63	181	4682
UE_Boys_04	30.75	10.95	26	56	181	4682
UE_Boys_05	25.96	16.06	19	63	181	4682
UE_Boys_06	31.33	14.6	25	65	181	4682
UE_Boys_07	29.96	16.06	23	67	181	4682
UE_Boys_08	39.49	14.96	33	74	181	4682
UE_Boys_09	46.95	9.12	43	68	181	4682
UE_Girls_03	57.44	8.53	37	61	181	4682
UE_Girls_04	56.18	6.75	40	59	181	4682
UE_Girls_05	65.7	10.3	41	70	181	4682
UE_Girls_06	70.41	11.02	44	75	181	4682
UE_Girls_07	70.18	6.75	54	73	181	4682
UE_Girls_08	76	9.59	53	80	181	4682
UE_Girls_09	78.63	5.69	65	81	181	4682

Table 2: Summary statistics for Meshblock data. Prices, primary school income decile (ID), and University Entrance (UE) scores are distinguished by year, since these have a time component.

Table 2 contains summarises our GIS variables across the whole region. Hence these give us an idea of the choice set faced by households, while Table 1 tells us where they actually chose to live. House prices have generally trended up over the period considered, although (as in many parts of the world) there was a considerable slow down in growth during the 2008-2009 (Global Financial Crisis) period, reflecting tightening credit conditions and investor uncertainty.

The respondents living locations are reasonably prepresentative of the region's residential areas, being a mix of vegetation coverages, but predominately sparse vegetation. Industrial areas are relatively uncommon, with Wellington's employment scene being dominated by service and government. School performance has generally trended up over the time period considered, particularly in regards to university entrance scores.

3.2 Conditional logit results

Table 3 contains the results of our conditional logit model for household choice.

Examining first the tradeoff between time and income, we see that households have a strong preference for shorter commutes. Time is measured in hours, while price is measured in increments of \$100 000. Income is measured on a scale of 0-1, where 1 is "\$100 000 plus" and 0 is no income, with intervals of 0.1 corresponding to buckets of \$10 000. We note that the relative values of the coefficient on $\log(Time)$ and the coefficient on $\log(Price)$ give us an idea of the percent of extra house price a household is willing to pay in order to reduce their commute time by one percent. The ratio of 0.86 suggests that a household is willing to pay 0.86% more for a house that reduces the household head's commute by 1%. Assuming a couple who both work in similar locations, and who face a one-hour commute to work and own a \$300 000 house, this would mean that they would be willing to pay (assuming a 5.2% rental yield, or 0.1% per week) \$ $300 \times 0.0086 =$ \$2.58 per week to reduce their commute by 1%. A one percent reduction in commute time would save the couple 0.01 hours per commute, and with two commuters there would be twenty journeys per week, for a total of 0.2 hours (12 minutes). A valuation of \$2.58 for 0.2 hours translates into a valuation of \$12.90 per hour, which, for low income individuals, does not seem unreasonable.

As incomes rise, the sensitivity to house price declines (until, for the highest incomes, we are unable to discern a sensitivity to price), but so does the sensitivity to time, possibly reflecting ease of affording monetary commuting costs (which are highly correlated with time costs). We note, however, that the time and income interaction terms are not statistically significant.

Driving and Public Transport (PT) use are less popular than Active Transport (AT; the baseline), all other things being equal. However, we note that in most cases, time costs are substantially greater for Active Transport, making it a less attractive proposition. Further, age (scaled in this regression by 100) has a negative effect on AT usage. Considering a 30 year old commuter, the age AT effect will counteract nearly half of the driving effect. Considering our examination of a 30 year old commuter, dividing the drive coefficient by the time coefficient suggests that if driving is $e^{(4.1344-0.3\times 6.3271)/1.2227} = 6.2$ times as fast as active transport, then this would make driving the dominant modal choice (all other things being equal). Given speed limits of 100km/h on motorways and 50km/h on suburban roads, this will generally make driving more attractive for this age group. Public transport is considerably less attractive, and the ratio grows to $e^{(6.1444-0.3\times 6.3271)/1.2227} = 32.2$ meaning

Variable	Coefficient	t-stat	Variable	Coefficient	t-stat
$\log(Time)$	-1.2227	-6.4635	$Bedrooms$	0.2921	2.0287
$\log(Time) * Inc$	0.4105	0.6854	$Bedrooms * Nkids$	0.0234	0.5679
$\log(Time) * Inc^2$	1.2282	0.2460	$PercNorth$	0.8623	4.8025
$\log(Price)$	-1.4147	-3.1846	$DenseMB$	1.1469	3.4682
$\log(Price) * Inc$	1.5444	2.8426	$DenseAU$	-2.6242	-5.6774
			$SparseMB$	0.4287	1.1762
$Drive$	-4.1344	-7.6306	$SparseAU$	-1.0694	-1.6529
PT	-6.1444	-11.2513	$CommMB$	-4.3325	-1.9614
$HiCar$	1.5555	4.0292	$CommAU$	1.7409	1.5016
			$IndustMB$	-8.4089	-3.6922
$WorkDT * Drive$	-2.3950	-11.7771	$IndustAU$	2.8874	8.1158
$Age * AT$	-6.3271	-5.8819			
$NoLic * Drive$	-2.6916	-5.0545	$GoodPrim$	-0.2110	-0.4359
$AT * Inc$	-4.6944	-2.8569	$GoodPrim * NK10$	-0.0705	-0.5179
$AT * Inc^2$	2.4469	0.1567	$GoodPrim * Income$	0.7124	1.1125
			$GoodSec$	0.5352	1.0281
$Female * Drive$	0.1663	0.6308	$GoodSec * NK1118$	-0.0210	-0.0828
$Female * PT$	0.6705	2.3297	$GoodSec * Income$	-1.4686	-2.1009
$Female * HiCar$	0.0444	0.1899	$Residual$	-0.1578	-0.4720
$Maori * Drive$	0.3850	0.9619			
$Maori * PT$	-1.0501	-1.5218			
$Maori * HiCar$	-0.7365	-1.7707			
$Pasifika * Drive$	1.1150	1.2751			
$Pasifika * PT$	1.7307	2.2044			
$Pasifika * HiCar$	-0.7959	-1.2360			
$Asian * Drive$	0.7830	1.3509			
$Asian * PT$	0.9226	1.9358			
$Asian * HiCar$	0.1175	0.2703			
$NCarLicn * HiCar$	-0.3096	-2.4680			
$NWorkers * HiCar$	-0.1937	-1.4069			
$Income * HiCar$	0.2850	0.6284			

Table 3: Conditional Logit model for household decisions of where to live, how to commute, and how many cars to own. This model is estimated as described in Section 1, using two-stage estimation of the household decisions, joint with the house pricing.

that Public Transport is generally preferred to active transport only at highway speeds; commuters may favour walking or cycling to congested suburban buses. As income rises, active transport also becomes less attractive, and although this marginal effect declines as income rises, it is still negative across the income distribution.

Working in the downtown area presents a significant disincentive to driving, as does lack of a driver's licence (driving is still possible in this case, since a commuter may car-pool). Gender effects fall principally on PT use: women are considerably more likely to use buses and trains than their male counterparts. Maori, Pasifika, and Asian survey participants are more likely to drive, but PT use varies across the groups, with Asian and Pasifika more likely to use PT than Europeans, while Maori are less likely.

High car ownership is a popular choice (coefficient of 1.5555; t-statistic of 4.0292), but this effect declines as the number of car licences and number of workers in a household increases. Clearly our definition of high car ownership (requiring one car or more per driver's licence in a household) means that more cars are needed to satisfy this condition as licence numbers grow. However, the two sets of statistics do suggest that households are more likely to achieve economies of scale with vehicle ownership as household size (in terms of drivers and workers) increases.

Turning to residential characteristics, we find that large numbers of bedrooms are generally desirable. Households have a strong preference for north-facing meshblocks, due to greater sunlight. Vegetation has an interesting effect on a neighbourhood. Dense or sparse vegetation in a meshblock (perhaps representing a park or leafy neighbourhood) is a good feature, but vegetation in the area unit is a negative feature. Our examination of the neighbourhoods with high *area unit* vegetation revealed that these are predominately city fringe areas, where vegetated land may be farm or forest land that is privately owned, and not useful to residents. Commercial and Industrial areas have an opposite effect. These areas are desirable at the Area Unit level (presenting opportunities for employment and/or easy shopping) but not at the Meshblock level (where they may be noisy, smelly, or attract unwelcome interlopers into the neighbourhood). Lastly, our school effects are quite muted. We define a "good" primary school as one which has decile 5 or more, and a "good" secondary school as one that has a University Entrance pass rate of 50% or more. We find that higher income housholds have a preference for "good" primary schools. In contrast, lower income households have a preference for "good" secondary schools, while higher income hosueholds have a weaker preference for "good" secondary schools. We attribute this to a higher propensity for households to send their children to private secondary schools, making quality of state secondary school moot for higher income households.

4 Conclusion

This paper develops a model for commuting, residential choice, and car ownership for the Greater Wellington Region. In particular, it evaluates sensitivity to house prices, commute times, and various amenities. It also outlines new methodologies for dealing with endogeneity, and implementation of additional GIS variables.

We see there as being considerable potential for extension of this work. On the first front, quantifying individual preferences for modal choice and residential location allows extrapolation to census information. Given knowledge about the interaction between house prices, commute times, and ammenities, it should then be possible to consider comparative statics analyses of potential changes to infrastructure. For example, would changes to the roading network result in substantial changes to prices and demographic characteristics of the city?

Secondly, we see potential to extend the model to explore dollar cost effects on commuting. By exploring the effect of petrol price changes, we might gain insights into the elasticity of demand for driving as a commute mode.

Lastly, it seems possible to explore individual journeys reported in the HTS in more detail. Exploration of the "chains" of activities produced by individuals may shed further light on the motivation behind the preferences revealed in this paper.

References

Bhat, C. and J. Guo (2007), A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels, *Transportation Research Part B* 41, 506–526.

Blundell, R. and J. Powell (2004), Endogeneity in Semi-Parametric Binary Response Models, *Review of Economic Studies* 71, 655–679.

Comber, A., C. Brunsdon, and E. Green (2008), Using a GIS-Based Network Analysis to Determine Urban Greenspace Accessibility for Different Ethnic and Religious Groups, *Landscape and Urban Planning* 86, 103–114.

Conway, D., C. Li, J. Wolch, C. Kahle, and M. Jerrett (2010), A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values, *Journal of Real Estate Finance and Economics* 41, 150–169.

Guevara, C. and M. Ben-Akiva (2006), Endogeneity in Residential Choice Models, *Transport Research Record* 1977, 60–66.

Helbich, M., A. Jochem, W. Mücke, and B. Höfle (2013), Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning, *Computers, Environment and Urban Systems* 39, 81–92.

Higgs, G., R. Fry, and M. Langford (2012), Investigating the Implications of Using Alternative GIS-Based Techniques to Measure Accessibility to Green Space, *Environment and Planning B: Planning and Design* 39, 326–343.

Kaufman, D. and N. Cloutier (2006), The Impact of Small Brownfields and Greenspaces on Residential Property Values, *Journal of Real Estate Finance and Economics* 33, 19–30.

Petrin, A. and K. Train (2010), A Control Function Approach to Endogeneity in Consumer Choice Models, *Journal of Marketing Research* 47, 3–13.

Salon, D. (2009), Neighborhoods, Cars, and Commuting in New York City: A Discrete Choice Approach, *Transportation Research Part A* 43, 180–196.