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

Estimating location choice models in Australia

Stuart Donovan[1, 2], Thomas de Graaf[2], and Henri L. F. de Groot[2]

[1] Veitch Lister Consulting, 5/200 Mary St, Brisbane City QLD 4000

[2] Department of Spatial Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands

Email for correspondence: s.b.donovan@vu.nl

Abstract

We present preliminary results from ongoing research that seeks to estimate the location choice model formulated in Ahlfeldt et al. (2015). Using data for six Australian cities with populations ranging from 200,000 to over 5 million, we find the effect of commuting costs on location choice varies significantly between cities and across specifications. Our preferred model yields effects that are approximately 20% larger than conventional specifications.

1. Introduction

A growing body of research considers the effect of transport on location choice, or "land use" (see, for example, Ahlfeldt et al., 2015; Allen and Arkolakis, 2014). Questions remain, however, over the estimation and transferability of model parameters. In this paper, we present preliminary results from ongoing research that seeks to estimate the spatial general equilibrium (SGE) model of location choice from Ahlfeldt et al. (2015), hereafter "ASRW". [1] In contrast to studies that focus on a single city, we use data for six Australian cities with populations ranging from 200,000 to over 5 million. [2] We find the effects of commuting costs on location choice vary between Australian cities and across model specifications, with our preferred model yielding effects approximately 20% larger than conventional specifications.

2. Model

Here, we summarise only the relevant parts of ASRW; readers are referred to the paper for full details. In ARSW, workers choose their home and work locations based on prices, such as rents and wages; amenities at home and work; and commuting costs. In spatial equilibrium, prices, amenities, and commuting costs adjust to leave workers indifferent between locations. Formally, the preferences, U_{ijo}, of worker o over home and work locations i and j is given by

$$U_{ijo} = z_{ijo} \frac{B_i}{d_{ij}} \left(\frac{c_{ijo}}{\beta}\right)^{\beta} \left(\frac{l_{ijo}}{1-\beta}\right)^{1-\beta} ,$$

where z_{ijo} is a random variable that denotes the idiosyncratic preferences of worker o for home and work locations i and j; B_i denotes the level of residential amenities; d_{ij} denotes the disutility of commuting between i and j; c_{ijo} denotes a composite consumption good; l_{ijo} denotes residential floor space; and $1-\beta$ denotes the share of housing in total consumption.

[1] Lennox and Sheard (2019) use a variant of ASRW to analyse urban transport improvements in Australia.

[2] We present results for Sydney (SYD), Melbourne (MEL), South East Queensland (SEQ), Perth (PER), Adelaide (ADL), and Hobart (HOB). Results for Launceston and Canberra will be added in the future.

We assume full-time workers supply one unit of labour and earn wage w_j, whereas the price of floorspace is given by q_i and the price of the composite good is the numeraire. ASRW assume z_{ijo} follows a Fréchet distribution given by $F(z_{ijo}) = \exp(-T_i E_j z_{ijo}^{-\varepsilon})$, where $T_i, E_j > 0$ are scale parameters for the average utility from living and working in i and j, respectively, and $\varepsilon > 1$ measures heterogeneity in worker's preferences over home-work locations. A smaller value for ε implies more heterogeneous preferences over locations, and vice versa.

Under these assumptions, ASRW integrate over z_{ijo} to derive the following equation to describe the probability, π_{ij}, that an individual worker chooses to live and work in i and j:

$$\pi_{ij} = \frac{T_i E_j \left(d_{ij} q_i^{1-\beta}\right)^{-\varepsilon} \left(B_i w_j\right)^{\varepsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s \left(d_{rs} q_r^{1-\beta}\right)^{-\varepsilon} \left(B_r w_s\right)^{\varepsilon}}.$$

Where q_i and w_j denote the rents paid and wages earned by workers. Like ARSW, we model the disutility of commuting using the exponential function $d_{ij} = \exp \kappa \tau_{ij}$, where κ is a parameter to be estimated and τ_{ij} denotes commuting costs, which—in our case—is either car travel-time or generalized costs. Below, we focus on estimating the parameters κ and ε.

3. Methodology

We adopt a recursive methodology that broadly follows ASRW with some departures. First, like ASRW, we re-formulate the location choice model as a gravity equation (Eq. 1)

$$n_{ij} = \exp\left(\delta_i + \delta_j - \nu \tau_{ij}\right) \text{ [Eq. 1].}$$

Where n_{ij} denotes the number of workers who live and work in i and j ; $\delta_i = \ln\left(T_i B_i^{\varepsilon} q_i^{-\varepsilon(1-\beta)}\right)$ and $\delta_j = \ln\left(E_j w_j^{\varepsilon}\right)$ denote individual effects that measure the relative attractiveness of home and work locations; and $\nu = \varepsilon\kappa$ denotes the semi-elasticity formed from the product of ε and κ. We estimate Eq. 1 for each city, yielding six values for ν. Whereas Eq. 1 is often estimated as a Poisson model, we instead estimate it as a negative binomial model, which has a more flexible specification for variance.

Second, with the estimates for the semi-elasticity ν, we follow ASRW and use the following labour market clearing condition (Eq. 2) to estimate the transformed wages, ω_j:

$$H_{M_j} = \sum_{i=1}^{S} \frac{\omega_j / \exp\left(\nu \tau_{ij}\right)}{\sum_{s=1}^{S} \omega_s / \exp\left(\nu \tau_{is}\right)} H_{R_i} \text{ [Eq. 2].}$$

Where H_{R_i} and H_{M_j} denote the number of people who live in i and work in j, respectively, and $\omega_j = E_j w_j^{\varepsilon}$ denotes the "transformed wages" for location j. In equilibrium, there exists a unique vector ω_j that satisfies Eq. 2 exactly. Given the estimated semi-elasticity, ν; observed commuting costs, τ_{ij}; and observed population and employment totals H_{M_j} and H_{R_i}, we use Newton-Raphson's method to solve Eq. 2 for ω_j.

Third, we use our estimates for $\ln \omega_j$ to estimate the worker heterogeneity parameter, ε:

$$\ln \omega_j = \ln E_j + \varepsilon \ln w_j \ \text{[Eq. 3]}.$$

Here, we depart from ASRW by observing $\ln \omega_j$, is defined by both the individual destination effects, $\delta_j = \ln(E_j w_j^{\varepsilon}) = \ln \omega_j$ estimated in Eq. 1 and the transformed wages, ω_j, estimated in Eq. 2. Put another way, Eq. 1 and Eq. 2 provide us with two independent sources of information on $\ln \omega_j$. In Eq. 3, we use both these estimates, along with data on observed wages, w_j, and individual effects for each location to capture the average workplace utility, $\ln E_j$. To understand whether estimates of ε derived from observed wages, $\ln w_j$, are sensitive to sorting, we also test a variant of Eq. 3 where we replace $\ln w_j$ with estimates of the "spatial wage", $\ln w^s$ (Combes et al., 2008). The spatial wage represents the component of wages that remains after controlling for differences in sectoral composition (one-digit), occupational composition (one-digit), education, age, and gender between locations.

Finally, we use Bayesian methods to estimate the regression models in Eq. 1 and Eq. 3, which generates distributions of parameter estimates for ν, δ_j, and ε.

4. Data

We use two sources of data. Our first source is the Australian Census 2016.[3] From the Census, we extract home-work (SA2) flows n_{ij}, for full-time workers that commute by car, public transport, or active modes. We also extract data on the average wage, w_j, at the place-of-work and observed characteristics to estimate the spatial wage, $\ln w^s$, such as education and industry sector. Our second source of data is travel demand models operated by Veitch Lister Consulting, which provide us with data on commuting costs, τ_{ij}. From these models, we extract two commuting cost measures between SA2s: Car travel-times in the AM peak, τ_{ij}^t , and the generalized costs of travel, τ_{ij}^c. Whereas τ_{ij}^t facilitates comparisons to the existing literature, τ_{ij}^c, provides a more comprehensive, multi-modal transport cost measure.[4]

To this raw data, we apply two filters: First, we remove observations associated with origins and destinations with no residents or jobs, respectively. Second, we remove observations for which the car travel-time, τ_{ij}^t, exceeds 240-minutes, which are likely to be associated with non-regular commutes, such as fly-in-fly-out workers. After applying these filters, we are left with 500,000 observations, that is, unique SA2-origin, SA2-destination pairs. In Figure 1, we present scatter plots of the data for each of the six cities, where we show the logarithm of commuters, $\log n_{ij}$, versus travel-time, τ_{ij}^t, on the vertical and horizontal axes, respectively. In all cases, we observe an approximately linear, negative relationship between travel-time and the logarithm of commuters, as implied by the functional form of Eq. 1.

[3] Census data confers both advantages and disadvantages. On the upside, the Census enjoys high response rates and avoids issues with sample selection. On the other hand, the Census only records mode choice on one day.

[4] More specifically, generalized costs represent a "log-sum" measure of the disutility of commuting between locations, considering both the monetary and non-monetary costs of travel by car, public transport, and active modes. As we have models for several time periods in each city, we use the weighted-average costs between origins and destinations across all periods, where weights are defined by commute trips in each period.

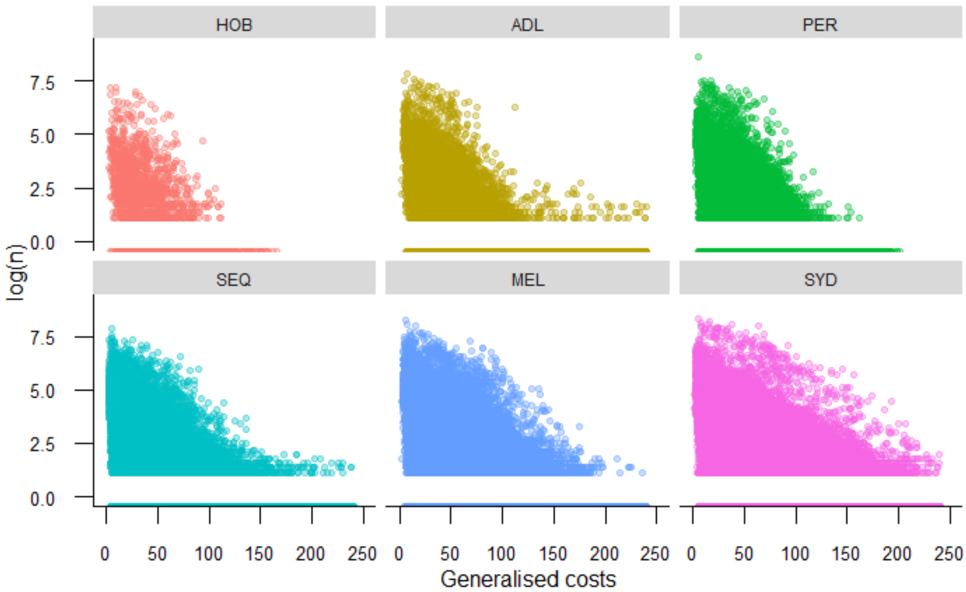

Figure 1: Scatter plots showing the logarithm of commuters, $\log n_{ij}$, versus travel-time, τ_{ij}^{t}, on the vertical and horizontal axes, respectively, for each of the six cities in our data, excluding observations $n_{ij} = 0$.

5. Results

Table 1 presents estimates of ν for three model specifications of Eq. 1 (rows) and city (columns). Model A is a conventional specification in which we use car travel-time, τ_{ij}^{t}, and treat δ_i and δ_j as fixed effects. In Model B, we replace car travel-time, τ_{ij}^{t}, with generalized costs, τ_{ij}^{c}. Then in Model C we use generalized costs, τ_{ij}^{c}, but treat δ_i and δ_j as random effects. To facilitate model comparisons, Table 1 presents standardised coefficients—where we divide τ_{ij}^{t} and τ_{ij}^{c} by their respective standard deviations prior to estimation.[5] In all models and cities, we find ν is precisely estimated with small standard errors.

Model	City					
	HOB	ADL	PER	SEQ	MEL	SYD
A	2.27 (0.04)	3.54 (0.02)	2.41 (0.01)	3.76 (0.01)	3.36 (0.01)	3.41 (0.01)
B	2.02 (0.04)	3.99 (0.02)	2.69 (0.01)	3.83 (0.01)	4.01 (0.01)	3.76 (0.01)
C	2.18 (0.04)	4.76 (0.02)	2.83 (0.01)	4.38 (0.01)	4.45 (0.01)	4.23 (0.01)

Table 1: Standardised coefficients for ν (c.f. Eq. 1); standard errors in brackets. Models described in text.

In terms of model performance, all models have R^2 values around 95%, although the PSIS-LOO information criterion indicates Model C has the best predictive performance. Comparing coefficients across models, we find Model C also returns the largest coefficients for all cities except Hobart. On average, coefficients for Model C are approximately 20% larger than those for Model A, where the latter is more commonly used in the literature. These are, in our view, economically meaningful differences. We also find significant differences in estimates between cities. The latter implies either (1) the underlying parameters, ε and κ, that make-up ν vary between cities or (2) endogeneity is biasing estimates of ν in ways that vary between cities, or both. Future work will consider endogeneity in more detail, although our preliminary results indicate it has only small implications for the magnitude of ν (NB: This further work will also investigate whether endogeneity in Eq. 1 affects estimates for δ_i and δ_j).

[5] On the original scale for τ_{ij}^{t} our estimates for ν in Model A are close to the 0.07 reported in ASRW.

With ν, we then use Eq. 2 to estimate transformed wages, ω_j, for each city. And finally, we estimate Eq. 3 by combining estimates of $\delta_j = \ln \omega_j$ from Eq. 1 with the estimates from Eq. 2. In Figure 2, we plot our estimates of $\ln \omega_j$ for Eq. 1 and Eq. 2 in the top and bottom panels versus raw income and spatial income in the left and right columns, respectively. Note that in each of these panels, the slope of the trend line provides an approximate indication of ε.

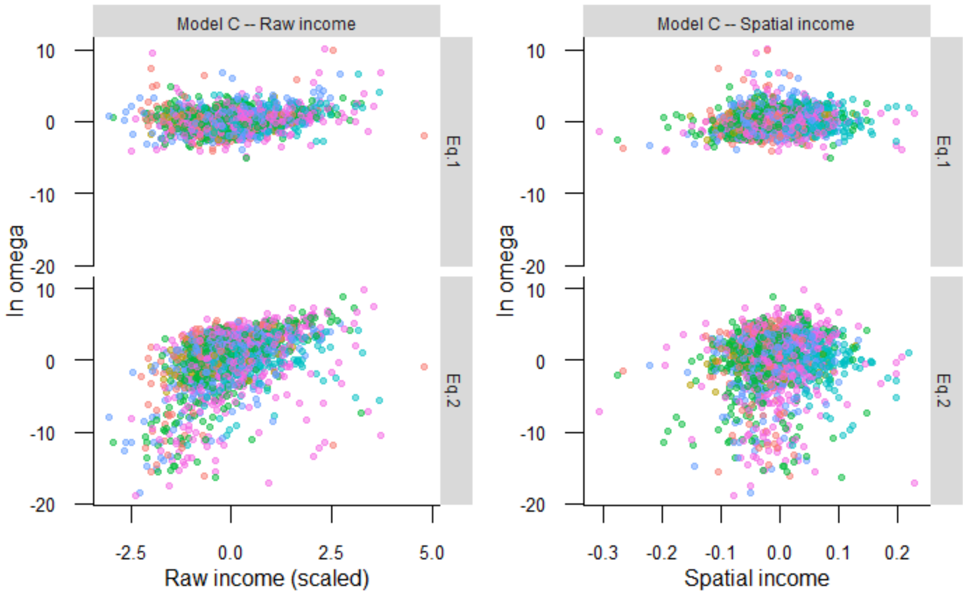

Figure 2: Scatter plots of $\ln \omega_j$ sourced from Eq. 1 (top row) and Eq. 2 (bottom row) on the vertical axes versus raw income $\ln w_j$ (left column) and spatial income $\ln w_j^s$ (right column) on the horizontal axes.

We prefer estimates of ε derived from Eq. 1 for both theoretical and empirical reasons. When estimating ε, however, we use the estimates of $\ln \omega_j$ from both the top and bottom panels of each column of Figure 2. As mentioned above, this provides us with two independent estimates of $\ln \omega_j$ per location, which in turn enables us to model average workplace utility, $\ln E_j$, by including individual location effects. We estimate Eq. 3 using raw income and spatial income, which yields estimates for ε of 3.58 (0.68) and 3.93 (1.52), respectively. Our preferred mid-point estimates for ε of 3.6-3.9 is smaller than that used in ASRW, which report $\varepsilon = 6.83$, but sits within the 1.33-5.62 range reported in Kazunobu et. al (2021).

In summary, using data for six Australian cities, we find the effects of commuting costs on location choice vary between cities and across model specifications. In further work, we aim to estimate models for Canberra and Launceston, refine the estimation of parameters, estimate additional model parameters, and consider implications for policy, such as local amenities, B_i.

References

Ahlfeldt, G.M., Redding, S.J., Sturm, D.M., Wolf, N., 2015. The Economics of Density: Evidence From the Berlin Wall. Econometrica 83, 2127–2189.

Allen, T., Arkolakis, C., 2014. Trade and the Topography of the Spatial Economy. Q. J. Econ. 129, 1085–1140.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! J. Urban Econ. 63, 723–742.

Kazunobu, H., Koster, H.R., Takatoshi, T., others, 2021. High-speed Rail and the Spatial Distribution of Economic Activity: Evidence from Japan’s Shinkansen.

Lennox, J., Sheard, N., 2019. Spatial general equilibrium analysis of a large urban rail project.