Should My Car Move or Should I? A Model of Residential and Commuting Choices

Christopher M. Clapp[*†‡]

November 4th, 2016

[Preliminary - please do not cite or quote without author's permission.]

Abstract

Urban communities across the country are implementing policies to address their ever increasing commuter congestion. These policies are relatively new and vary from city to city, so not much is known about their full effects. In order to evaluate different congestion reduction policies, I develop a discrete choice structural model of the joint decision of family residence and individual commuting mode, as well as the individual's resulting commute time, given the characteristics of the housing market and transportation system. I estimate the model using individual-level, restricted-access data from the American Community Survey (ACS) and a unique dataset of individual commute options and characteristics that I create using geographic information system (GIS) network analysis. I use model estimates to simulate the effects of transportation policies that alter the financial and time costs of commuting such as congestion pricing schemes, fuel or carbon taxes, and increased parking fees.

JEL Codes: Q52, R21, R41, and R48

Keywords: Residential Location, Travel Mode Choice, Congestion Pricing, Discrete Choice Analysis, Geographic Information Systems

[*]Address: Florida State University; Department of Economics; 113 Collegiate Loop; Tallahassee, FL 32306-2180; USA. Email: cclapp@fsu.edu. Phone: (850) 645-0756.

[†]This work has greatly benefited from conversations with Steven Stern, Leora Friedberg, Amalia Miller, Wayne Roy Gayle, Edgar Olsen, André de Palma, and numerous graduate students at U.Va. I would also like to thank Patrick Bayer, Bert Grider, and the staff of the U.S. Census Bureau's Center for Economic Studies (CES) for assistance with the restricted-access data and insightful comments. Finally, I would like to thank Chris Gist and Kelly Johnston for guidance on the use of ArcGIS to conduct my GIS network analysis, and Katherine Holcomb for assistance with the use of Fortran for my estimation program. All remaining errors are my own.

Part I
Introduction

Traffic jams are more than just a minor annoyance. American automobile commuters wasted an estimated 3.1 billion gallons of fuel and lost 6.9 billion hours because of congestion in 2014, a cost estimated at $160 billion (Schrank et al. 2015). Worse, congestion is not improving. The average annual congestion delay has more than doubled since 1982, the first year for which data is available. The social welfare cost of congestion is likely even greater than these estimates due to losses from uncertainty over commute times and congestion-induced increases in global and local pollution, traffic accidents, and noise.[1] Urban planners traditionally attempt to reduce congestion by increasing capacity: either by expanding roadways or public transit systems. Yet, Duranton and Turner (2011) find that building an additional kilometer of roadway leads to a one-to-one increase in mean daily vehicle kilometers traveled. They also show that the supply of mass transit alternatives has no effect on vehicle kilometers traveled. In other words, the most prevalent policy instruments for reducing congestion do not appear to have their intended effect.

As congestion continues to increase, communities across the country are looking at new congestion pricing policies that place a monetary cost on travel when and where congestion is greatest.[2] These policies can reduce congestion by internalizing externalities (Parry et al., 2007); yet, widespread use of such policies has been difficult for policymakers to implement due to constituent concerns that the policies are regressive in nature and fears that travelers will face increased financial costs without offsetting time savings. These priors persist despite examples of policies that have proven to be both successful and popular in a handful of cities around the world (Leape 2006, Eliasson 2014). Determining a priori whether these, and other, common voter misgivings about congestion reduction policies are warranted is a daunting task for two reasons. First, doing so requires predicting the impact of different policies on the well-being of very heterogeneous commuters in an interconnected transportation system. Second, congestion reduction policies affect far more than just how people commute. For instance, Baum-Snow and Kahn (2000), Bento et al. (2005), and Duranton and Turner (2011) all find evidence that commuting and residential location choices are inextricably tied. Housing choice influences the options in and characteristics of an individual's commuting choice set, so simple models of commuting decisions alone do not adequately capture behavioral responses. This work seeks to address these issues and inform the discussion of a burgeoning set of policy alternatives by simulating the effects of different congestion reduction policies.

To do so, I develop a structural model of an individual's commuting mode and the key interrelated decision affected by the policies of interest, residential choice. I make three key contributions to the literature. First, I deal with the endogeneity of residential choice in models of commuting method by explicitly modeling both residential and commuting choices together. Failure to ade-

[1] See Parry et al. (2007) for a full accounting of automotive externalities

[2] These policies include cordon charges that impose a fee on drivers who travel within or into a congested area and variably priced, managed lanes that prevent congestion by charging an adjustable access toll (Lewis 2008).

quately address this connection would result in biased coefficient estimates.

Second, I estimate my model using the 2000-2009 restricted-access versions of the ACS for the Washington, D.C. metropolitan area. These rich individual-level datasets allow me to include numerous unobserved heterogeneity terms which strengthen the validity of my results relative to more aggregate analyses which are often undertaken. I use precise information on where individuals live and work from the non-public versions of the ACS. The restricted-use geographic location information, along with data I have assembled on the structure of the transportation network allows me to map each individual's optimal commute from each home and by each commuting method in the choice set. To do this, I use geographic information system (GIS) network analysis. The mappings allow me to create a unique dataset of individual commute options and characteristics that I use to estimate the trade-offs that individuals make among consumption, housing amenities, and leisure when choosing a home and commuting mode pair. This allows me to improve upon the residential choice literature where state of the art analyses use controls for the role of commuting costs in housing decisions with either neighborhood aggregate commute times (Bajari and Kahn 2005, 2008) or as-the-crow-flies distances between home and work (Bayer and McMillan 2012).[3]

Finally, I am in the process of using model estimates to simulate the full set of effects of transportation policies that alter the financial and time costs of commuting on the joint distribution of residential housing and commuting methods. These policies include congestion pricing schemes, fuel or carbon taxes, and increased parking fees.

The next section provides a review of the related literature. Section III describes my theoretical model. I detail the data used in Section IV. Sections V and VI explain my estimation strategy and results. The results section in this paper is incomplete, but I outline the tasks that will be completed as part of the published version of the project. Finally, I offer conclusions in Section VII.

Part II

Literature Review

My research draws from three distinct literatures: transportation, residential location choice, and congestion pricing. I begin by discussing insights from the former areas, as well as highlighting ways in which my work advances the given literature. Then, I conclude by providing background on congestion reduction methods.

[3] Langer and Winston (2008), who are also interested in the effects of congestion reduction policies, ask, "How can one estimate the economic effects of road pricing while accounting for its impact on land use?" They posit a methodology similar to the one I develop, but note that it is an ambitious undertaking in explaining their decision to use hedonic methods to answer the question, by saying, "a disaggregate approach for a metropolitan area would model the determinants of a commuter's choice of mode of transportation, departure time, destination, route and residential location and simulate how those choices change in response to an efficient congestion toll. Unfortunately, the data and modeling requirements of a disaggregate approach-especially in determining a commuter's residential location alternatives and their attributes-are formidable." My work attempts to make progress on the formidable task they describe.

1 Transportation

There has been a great deal of research on what can broadly be categorized as travel demand analysis, and Small and Verhoef (2007) provide a comprehensive overview of the issues that must be addressed and the methods that are used. Key among those issues, in order to estimate how individuals commute, one needs to find a way to measure the alternative-specific attributes of commutes that an individual did not choose. As Small and Verhoef explain, there are two options: either use values reported by individuals in the survey or use engineered values produced from network analysis. Each has shortcomings. The former may be biased because individuals do not know much about the options that they do not choose or because they misreport so as to reinforce the option they do choose. The latter are costly to calculate and are not always accurate. ACS data reports commute times for chosen options only, so I calculate engineered values using GIS network analysis. Although computationally expensive to calculate, results in Section 10 show that they do explain some of the variation in reported commute times, conditional on commuting by the given method.

I use the engineered commute times I calculate as an input to a disaggregate model of commuting modes based on the random utility maximization (RUM) model developed by McFadden (1974).[4] The basic identically and independently distributed (iid) multinomial probit (MNP) and logit (MNL) versions of this model suffer from the well known independence from irrelevant alternatives (IIA) problem, so subsequent research has relaxed this assumption, commonly with nested MNL or mixed-multinomial logit (MMNL) models (McFadden 2001). My structural model also relaxes these assumptions, as well as the assumption that an individual, choice specific error enters utility linearly, as there is no strong economic justification for this specification.

Finally, Small and Verhoef (2007) argue that the endogeneity of travel characteristics is not a great concern when using disaggregate data because researchers can make the assumption that individuals take those characteristics as given when making travel related decisions. This is true to the extent that individuals have no control over the characteristics of their commute. However, Baum-Snow and Kahn (2000) find evidence that some of the increase in system travelers after the expansion of mass transit systems can be attributed to individuals who move to take advantage of the infrastructure improvements, and Duranton and Turner (2009) indicate that individuals have the same response to the expansion of roadways. Although individuals cannot control, for instance, how fast traffic flows on a given road or where a given subway train stops, they can control which road or which subway line they take by choosing where they live relative to where they work.[5] Additionally, Bento et al. (2005) find evidence of a relationship in the other direction: measures of urban spatial structure have small but significant effects on travel demand. They provide a thorough discussion of the bias that is caused by failure to adequately address this connection. In order to address the biases introduced by the interdependence of commuting characteristics and residential

[4] McFadden (2001) provides a historical survey of the methodology of individual travel demand analysis using RUM models, and Train (2009) is an excellent resource for practitioners.

[5] To be precise, work location also influences commuting characteristics. See Appendix B.3 for a discussion of the implications of not including work location decisions in my model.

location, I jointly model both residential choice and commuting decisions.

2 Residential Choice

Early, theoretical work on location decisions is characterized by the assumption that all individuals commute to the same central business district, and a land-rent gradient develops (see, Alonso 1964, Mills 1967, and Muth 1969). Evidence of this gradient can be seen in current empirical research. Bajari and Kahn (2005, 2008) model residential location decisions using a three-step estimation process based on hedonic estimation of home prices. The latter work explicitly controls for commuting costs with the average commute time of individuals who live in the given home's Census tract. They find that willingness to pay to reduce commuting time is slightly less than the household owner's hourly wage at the margin.[6] Langer and Winston (2008), who also use hedonic methods, calculate a marginal willingness to pay of roughly half the average household wage when using the average commute time in the household to measure commuting costs. While these aggregate measures of commuting are useful in hedonic settings where the value of the home is determined by market forces (not just an individual's valuation), they are less satisfactory in a model of individual outcomes. I model the actual commuting options and characteristics that individuals face when choosing a home. I also relax the implicit assumption that all commuters travel to the same area for work and model cities as aspatial urban areas instead of monocentric ones.

Bayer and coauthors have pioneered an alternative way to model residential location using restricted-access Census microdata to explore topics ranging from segregation in housing markets (Bayer et al. 2004) to labor market hiring networks among neighbors (Bayer et al. 2008). In general, these works estimate equilibrium models of residential choice using household data and the differentiated products methods of Berry et al. (1995).[7] They model differences in household preferences for residential locations, conditional on work location, but focus on the implications of Tiebout (1956) sorting in housing markets, not commuting decisions. They control for the influence of commuting in residential decisions with the as-the-crow-flies distance to the head of household's job. Bayer and McMillan (2012) find, for instance, that households are willing to pay about \$50 per month to reduce daily commutes by one mile. I improve on their methodology by more accurately modeling the commute faced by individuals in the household. Specifically, I model the duration and mode of the commute, and I allow for heterogeneity in preferences over commuting methods. They also find that their commuting estimates are sensitive to controls for unobserved neighborhood quality, so I develop a flexible way to incorporate neighborhood effects into my model that is outlined in Section 12.1.2.

The earliest empirical models of residential choice were developed in the late 1970s (Lerman 1976, McFadden 1978) based on the RUM model. While Lerman (1976) also incorporates com-

[6]They note that this estimate is greater than estimates of roughly half the hourly wage commonly found in the transportation literature.

[7]See Ioannides (2012), Chapter 4 for a review.

muting decisions (as well as automobile ownership), I know of only one recent paper that models the joint decision of residential location and commuting mode using individual level data. Vega and Reynolds-Feighan (2009) use GIS network analysis to augment individual-level data to estimate a cross-nested logit (CNL) model of the joint residential location and commuting decision. They find that commuters have heterogeneous mode choice responses to policies that increase their travel costs and that congestion policies are likely to have an effect on residential location decisions. Although Vega and Reynolds-Feighan model commutes using GIS techniques for both automotive and mass transit options, they aggregate all of the work locations in their city of analysis (Dublin, Ireland) to four employment centers, resulting in a loss of precision. I improve on their methodology by more accurately controlling for the commute characteristics individuals face, as well as by allowing for a more flexible error structure in my model.[8]

3 Congestion Pricing

Economists have long advocated for use fees that internalize congestion externalities and improve welfare.[9] Lindsey (2006) provides a comprehensive survey of the theoretical literature on road pricing dating back to Adam Smith, but congestion pricing policies have only more recently begun being implemented and are still not widespread.[10] Lewis (2008) provides an overview of the various forms of congestion pricing policies which I summarize in Table 1.

Table 1: Congestion Pricing Policies

Type	Definition	Examples
Area Wide	Charges based on congestion level on all roads	None
Variable Roadway	Tolls include rush hour fees for particular roads	NJ Turnpike
Managed Lanes	Variable tolls for separated lanes within a highway	I-15 & SR-91 (CA)
Cordon	Fee to drive within or into a congested area	London
Zonal	Cordon charging with adjacent charging zones	Trials in Europe

Source: Table created by the author using information from Lewis (2008).

He argues for the effectiveness of congestion pricing policies with some impressive measures. The cordon charge introduced in London, England in 2003 reduced traffic in the cordon by 20 percent, increased traffic speeds by 37 percent, and raised more than $100 million in net revenues that were used to improve the city's mass transit system. Leape (2006) reports that the London

[8]My current work focuses on individual decision-makers and results are based on households with a single adult commuter. Future research will also account for household bargaining over commuting characteristics between spouses. This will allow me to add cohabiting couples to the model, instead of just relying on single individuals for estimation as is the case in both this work and Vega and Reynolds-Feighan (2009).

[9]Parry et al. (2007) discuss the externalities associated with automotive travel and the policies, ranging from fuel taxes to congestion pricing, that can be used to address those externalities. The discussion is both in terms of efficacy and political feasibility.

[10]See https://ceprofs.civil.tamu.edu/mburris/pricing.htm for a list of all instances of congestion pricing in practice today. At present, there are less than 50 (broadly defined) examples of congestion pricing on roads around the world.

cordon charge has been such a success that there have been discussions of a nationwide congestion pricing policy. In the United States, managed lanes on SR-91 in California had an average speed of 60 miles per hour during peak hours while congestion in the untolled lanes reduced speeds to under 20 miles per hour.[11]

Small et al. (2005, 2006) perform a thorough analysis of the effects of the congestion pricing mechanism used on SR-91, finding that it does improve motorist welfare due to significant heterogeneity in traveler preferences. This occurs because low-value-of-time commuters are displaced by high-value-of-time commuters who reap large benefits. However, the available data prevents the authors from modeling mode choice, residential location, or time of travel, all of which can be varied by commuters in the long run. My model addresses these concerns, while allowing for a robust set of unobserved heterogeneity parameters their work suggests are important.

Part III

Model

This section outlines a structural model of residential choice and commuting method that I estimate using Census microdata. The structure allows me to determine the relative importance of housing and neighborhood characteristics on residential choice, including distance to place of employment and access to commuting options. I allow for heterogeneous preferences for those characteristics as well as for commuting methods. I detail the model as it pertains to a single, adult decision-maker.[12] Finally, I explain how the model addresses the impact that dependent children have on the behaviors of their parents.

While I advance the literature by treating the choice of residential location and commuting mode as joint in a model with as much geographic detail as I have, I must nevertheless take other decisions as fixed in order to keep the model tractable. I assume that an individual takes her city of residence, family structure, vehicle ownership, and employment as given when deciding among transportation options and residential choices. Additionally, I assume that the locations and hours of employers and schools are independent of residential choices and transportation options. All of these assumptions have the potential to bias my results, although to varying degrees. I discuss the implications of these assumptions in Section B.3.

4 Single Person Household

The simplest type of family is that of an individual choosing where she alone will reside and how she will commute. I build from the standard labor-leisure framework. An individual has prefer-

[11] See Anas and Lindsey (2011) for more information on the effects of several major congestion pricing programs.

[12] I leave the case of a family containing two, adult decision-makers who bargain with one another for future research. Households populated by roommates are not considered.

ences over consumption and leisure and faces both a budget and time constraint. Consumption is defined over a composite good and housing amenities, and leisure takes the form of either time spent away from work or of some fraction of time spent commuting.

4.1 Preferences

I define a market (indexed by m) at the metropolitan level and assume that jobs (j) have characteristics that include wages, hours, and location. Given a fixed market and job, an individual (i) is faced with the decision of which house to live in (h) and by what method to commute (k). Preferences are defined over composite consumption (c_{ihk}), housing amenities ($\tilde{H}_{ih}$), and leisure ($\tilde{\ell}_{ihk}$) and represented by a utility function as

$$U\left(c_{ihk}, \tilde{H}_{ih}, \tilde{\ell}_{ihk}\right).$$

The aggregate consumption good, c_{ihk}, includes all non-housing consumption and savings.

As in Bayer et al. (2005), the individual derives utility from many different housing amenities, including characteristics of both the house and the neighborhood. In order to include a rich set of housing characteristics but still keep the utility function tractable, I define $\tilde{H}_{ih}$ as a function of observable housing and neighborhood characteristics (H_{ih}) and unobservable characteristics (ε_{ih}),

$$\tilde{H}_{ih} = \exp\left(H_{ih}\gamma^{H} + \varepsilon_{ih}\right).$$

The $\exp(\cdot)$ ensures that the utility function can be evaluated.[13] The observable characteristics, H_{ih}, are allowed to vary over both i and h in order to allow for interactions between individual and home-specific observables, but variation over individuals is not necessary for the identification of γ^{H}. The error term, ε_{ih}, is necessary to explain cases where an individual chooses to live in a home that is observationally inferior to other homes in her feasible choice set. It can account not only for unobserved characteristics of the home, but also for search and moving costs that might lock an individual into a given home, but that are not explicitly modeled. It is known to the agent but not to the econometrician, thus providing a source of unobserved heterogeneity in the model.

Following McFadden (2001), non-work time has two components,

$$\tilde{\ell}_{ihk} = \ell_{ihk} + (1 - \lambda_{ik})\, t_{ihk}.$$

The ℓ_{ihk} term represents pure leisure. Time spent by individual i commuting between home h and job j by method k (t_{ihk}) may contain a leisure component that is known to the agent but not the econometrician.[14] This component accounts for heterogeneity in preferences for commuting methods to explain cases where an individual chooses to commute by a method that is more costly, both in terms of time and money, than other feasible methods.[15] It is measured by the random variable

[13]In a subsequent section, I specify the utility function with a log transformed Cobb-Douglas functional form. The $\exp(\cdot)$ ensures that $\tilde{H}_{ih} > 0$ so that $\ln\left(\tilde{H}_{ih}\right)$ can always be evaluated.

[14]Note that I drop the j subscript in all variables that vary over i, as jobs are taken as fixed for a given individual.

[15]That heterogeneity in preferences for commuting methods affects utility through leisure is an assumption. This

λ_{ik}, which is bounded from below at 0 and varies over individuals and methods of commuting. As such, if $\lambda_{ik} = 0$, time spent commuting by method k is a perfect substitute for pure leisure. Note that if commuting by method k is stressful and work-like, $\lambda_{ik} \cong 1$. A value of $\lambda_{ik} > 1$ means that the individual views commuting to be less enjoyable than work.

There is nothing in economic theory that requires a lower bound on λ_{ik}, but $\lambda_{ik} < 0$ does not seem plausible. A value of $\lambda_{ik} < 0$ would mean that the individual would rather commute than engage in general leisure activities. Since traveling by method k is a feasible leisure activity, I restrict λ_{ik} to prevent nonsensical preferences.[16]

4.2 Prices

Individual i takes as given several prices in her market. The price of the aggregate consumption good varies by metropolitan area. A local cost-of-living index, denoted as p^c_m, is used to measure this variation. The opportunity cost of owning or renting a home is imputed as in Bayer et al. (2007) and is represented as p^H_h.[17] I do not observe savings or wealth, nor does my data allow for a dynamic model, so converting housing stock expenses into flow opportunity costs is necessary, given that a savings motive does not drive housing choice in my model. The average pecuniary cost per mile of commuting via method k in market m is denoted as p^d_{mk}, where the d superscript denotes distance.[18]

In the data, there are 12 reported methods of commuting. These methods are condensed to the most relevant options in Table 2, with associated per mile commuting costs.

Household automobile ownership is observed only as the number of vehicles available, but not make, model, or year of those vehicles, so I use an average measure of miles per gallon (MPG) to determine the price of commuting by automobile.[19] The number of people in the carpooling option is denoted by N^{pool}. The $\bar{p}_{mk}$ prices are the average fare per mile for the given system in

specification is useful because it allows the preference to vary with the duration of the commute. An individual may have an extreme dislike for driving, but may opt to drive if a short commute minimizes the displeasure. This specification is less desirable if the costs or benefits of a given method of commuting are not variable. For instance, if an individual prefers to drive because of the flexibility it allows in running errands after work.

[16]This restriction is supported by the the time use literature. Krueger et al. (2008) provide comparisons of how people felt while engaging in different activities. Unsurprisingly, their results indicate that individuals prefer most leisure activities to commuting. Additionally, their results show that commuting and working rank as almost equally unenjoyable activities, with their ordinal rankings varying by survey methodology.

[17]For notational clarity, I capitalize the "H" superscript that serves as a label for the price to avoid confusion with the "h" subscript that serves as an index.

[18]Fixed costs, such as parking fees and tolls, are assuredly important components of commuting decisions, but I do not observe these costs in the data. The former is not reported by individuals and the latter depends on the exact commuting route, which I do not observe.

[19]I assume 20 MPG based on Bureau of Transportation Statistics "Average Fuel Efficiency of U.S. Light Duty Vehicles" figures (http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics /html/table_04_23.html). Since not all vehicles have the same fuel efficiency, it would be preferable to specify MPG as the sum of the mean MPG of the automotive fleet in the given year and an individual specific error. This would allow me to integrate over the distribution of the error in order to obtain a more accurate measure of automotive commuting costs, as well as allow for correlation with other errors in the model. Due to the added computational costs, this is left for future research.

Table 2: Commuting Methods and Costs

Method (k)	Pecuniary Cost Per Mile
Automobile	$p^d_{m,\text{auto}} = \frac{p_{\text{gas}}}{MPG}$
Carpool	$p^d_{m,\text{pool}} = \frac{p_{m,\text{auto}}}{N^{pool}}$
Bus	$p^d_{m,bus} = \bar{p}_{\text{m,bus}}$
Streetcar	$p^d_{m,streetcar} = \bar{p}_{\text{m},streetcar}$
Subway	$p^d_{m,subway} = \bar{p}_{\text{m},subway}$
Rail	$p^d_{m,rail} = \bar{p}_{\text{m,rail}}$
Walk	$p^d_{m,other} = 0$

metropolitan area m.

4.3 Constraints

Individual i faces both a budget constraint and a time constraint. To represent expenditures, I first define N^H_m and N^K_m as the number of homes and commuting methods in market m. I then define an $N^H_m \times 1$ vector, I_i, whose hth element is 1 if the individual lives in home h and 0 otherwise. Next, I define d_{ihk} as the distance between house h and job j that individual i travels by commuting method k. I pack those distances into an $N^K_m \times 1$ vector of commuting distances traveled by individual i from house h by each commuting method, d_{ih}.[20] The budget constraint is defined as

$$p^c_m c_{ihk} + \underset{(1\times N^H_m)}{p^H} \underset{(N^H_m\times 1)}{I_i} + \underset{(1\times N^K_m)}{p^d_m} \underset{(N^K_m\times 1)}{d_{\text{ih}}} = w_{\text{i}} L_{\text{i}},$$

where w_{i} is individual i's wage, and L_{i} is the individual's time spent at work. Sample selection criteria guarantee that all individuals are employed, and wages and work hours are taken as fixed.

I denote total time as T and the commuting time by method k as t_{ihk}. Individual i's time constraint is

$$\ell_{ihk} + t_{ihk} + L_{\text{i}} = T.$$

The commuting time by method k (t_{ihk}) is treated as a function of a linear index of the characteristics of the commute ($K_{ihk}\gamma^K_k$) and a measurement error term (e_{ihk}) due to the econometrician's uncertainty about the exact route the agent takes, traffic patterns, etc.[21] It is written as

$$t_{ihk} = \exp\left(K_{ihk}\gamma^K_k + e_{ihk}\right), \tag{1}$$

where the $\exp(\cdot)$ ensures that time spent commuting is positive. Note that random, temporary shocks (e.g., accidents, weather, construction) do not affect the agents' long term commuting decisions.

[20]Note that the kth element of d_{ih} will be 0 for individuals who do not commute by the kth method.

[21]This specification is necessary because the data reports commute times for chosen options only. I use characteristics of the commute calculated using GIS network analysis to impute unobserved commute times based estimates of γ^K recovered from observed commute times.

4.4 Parameterization

I define the utility function with a Cobb-Douglas functional form and make the familiar natural log transformation, which results in

$$U\left(c_{ihk}, \tilde{H}_{ih}, \tilde{\ell}_{ihk}\right) = \alpha_i^c \ln\left(c_{ihk}\right) + \alpha_i^H \left(H_{ih}\gamma^H + \varepsilon_{ih}\right) + \alpha_i^\ell \ln\left(\ell_{ihk} + \left(1 - \lambda_{ik}\right) t_{ihk}\right),$$

where α_i^c, α_i^H, and α_i^ℓ are taste parameters over composite consumption, housing amenities, and leisure. I normalize α_i^c to 1 and γ_0^H to 0 to ensure identification. The other parameters, α_i^H and α_i^ℓ, are allowed to vary with observable characteristics (X_i) of the individual and contain error terms to capture unobserved heterogeneity in preferences. The taste parameters are defined as[22]

$$\begin{aligned}
\alpha_i^c &= 1, \\
\alpha_i^H &= \exp\left(X_\mathrm{i}\beta^H + \mu_i\right), \\
\alpha_i^\ell &= \exp\left(X_\mathrm{i}\beta^\ell + u_i\right).
\end{aligned}$$

4.5 Choice Problem

Taking labor market decisions, job characteristics, and vehicle ownership as given, the full choice set is a residence and a method of commuting. By choosing a residence, the individual determines the characteristics of both her home and commute options. The joint choice of a residential location and a particular method of commuting determine the individual's consumption and allocation of time. The former is uniquely determined by the budget constraint, since there is no saving in the model. Similarly, since hours of labor are taken as given, the time constraint determines leisure. In summary, an agent in the model faces the unconstrained choice problem

$$\begin{aligned}
\max_{h_i, k_i} U\left(c_{ihk}, \tilde{H}_{ih}, \tilde{\ell}_{ihk}\right) &= \ln\left(\frac{w_\mathrm{i}L_\mathrm{i} - p^H I_\mathrm{i} - p_m^d d_\mathrm{ih}}{p_m^c}\right) + \exp\left(X_\mathrm{i}\beta^H + \mu_i\right)\left(H_{ih}\gamma^H + \varepsilon_{ih}\right) \\
&+ \exp\left(X_\mathrm{i}\beta^\ell + u_i\right) \ln\left(T - L_\mathrm{i} - \lambda_{ik}\exp\left(K_{ihk}\gamma_k^K + e_{ihk}\right)\right). \qquad (2)
\end{aligned}$$

5 Households with Children

Children are an important factor in the housing and commuting decisions of their parents. In order to capture the effect that children have on housing and commuting decisions, I include the presence and characteristics of children and the interaction of these terms with key housing and neighborhood characteristics. For instance, the interaction of local school quality with the presence

[22]Note that the exponential form guarantees that the utility parameters will be positive, ensuring that "goods are good."

of children in the household will help control for a parental desire to send their children to high quality schools. However, I do not explicitly model the commuting behavior of children in the household.[23]

Part IV

Data

This section outlines the main data sources I use and how they are linked. From Equation 2, it can be seen that to estimate my model I need to observe three outcomes: housing choice (h_i), commute method (k_i), and commute time (t_{ihk}). I also need data on housing characteristics (H_{ih}), commute characteristics (K_{ihk}), and individual characteristics (X_i). Additionally, in order to recover composite consumption, I need data on the prices of composite consumption (p_m^c), homes (p^H) and commuting methods (p_m^d). No single dataset contains all of this information. In order to construct a dataset that allows me to estimate my model, I combine data from the U.S. Census Bureau's ACS and the U.S. Department of Transportation's National Transportation Atlas Database (NTAD) using GIS mapping software. I also augment that dataset with pricing information from various sources.

The main dataset my analysis is built on is the restricted-access Census microdata versions of the 2000 - 2009 ACS. [24] The ACS contains information similar to the Decennial Census Long Form Questionnaire that it replaced after the 2000 Census. It is an annual sample of one in 40 households in the country.[25] The Census Bureau first began producing ACS data in 2000 to test the survey and officially began producing the survey in 2005, so my data is a repeated cross-section.

There are two key features of ACS data that are important for my research. First, ACS surveys include questions on place of residence, primary commuting method, and commuting duration, as well as individual and household characteristics and linkages that allow the identification of the relationship between members of a household. Second, while the ACS does not contain a great deal of information about individual commute characteristics, it does report the daily commute time and includes information about the place of residence and place of work that allow me to augment the commuting data. The restricted versions of these datasets allow me to identify both the home and work locations of each individual down to the Census block, which provides the geographic precision necessary to calculate unobserved commute characteristics in a meaningful way. I detail the former feature first, then provide more detail on the latter in subsequent sections. I conclude the section by discussing the additional price data I use.

[23]Doing so is a feasible extension of this line of research that requires modifying the the GIS network analysis I perform for adult job locations to child school locations. I leave this exercise for future research.

[24]I discuss the shortcomings of the publicly available data in Section 9.

[25]For reference, every decade the Long Form sampled one in 6 households. See http://www.census.gov/acs/www/Downloads/handbooks/ACSPUMS.pdf for more information.

6 Sample Selection

I begin by defining markets (m) using the the Office of Management and Budget's (OMB) definitions of metropolitan areas. The OMB creates these designations for use by federal agencies in statistical analysis. Metropolitan areas are defined as central urban areas and any adjacent counties that have "a high degree of social and economic integration (as measured by commuting to work) with the urban core."[26] The OMB defines Combined Statistical Areas (CSA) to represent contiguous urban areas (ie, Washington, DC and Baltimore, MD) and Core Based Statistical Areas (CBSA) to represent central (ie, Pittsburgh, PA) or component (ie, Washington, DC) cities.[27]

I restrict all data to the "Washington-Arlington-Alexandria, DC-VA-MD-WV" CBSA (hereafter refereed to as the DC CBSA) in order to keep the estimation tractable.[28] I use this definition of the market for all years of the data even though it was created in 2003 to avoid using a varying definition of the market each year. I choose this CBSA for several reasons. First, it has the most automotive commuter congestion in the nation according to Schrank et al. (2015), so there is a need for the policy analysis I perform. Second, there is a robust mass transit system in the Washington, DC area that allows individuals to respond to a given policy change in multiple ways. This both increases the need for the simulations I perform and allows for the analysis of numerous policy options. Finally, the District offers a great deal of geographic information that is not available nationally which is accessible through the District of Columbia Geographic Information System (DC GIS). Specifically, although the NTAD contains geographic location information for rail systems, it does not have comparable information for bus routes that DC GIS makes available for Washington Metropolitan Area Transit Authority (WMATA) bus lines and stops.

I restrict the sample based on observable characteristics at both the household and individual levels. First, I drop households based on housing unit characteristics that indicate that the residence may not be the family's primary home or that the full financial costs of the home are not fully captured by the questions asked in the ACS. Second, I restrict the sample based on household characteristics that indicate that the household's income is in the tails of the income distribution or based on relationships in the household that indicate that the household bargaining process is too complex to model without additional information (for instance, a parent and adult children living in the same household).[29] Third, based on individual characteristics, I drop all households

[26] See http://www.census.gov/population/metro/ for more information.

[27] In 2003, the OMB updated the names and definitions of core metropolitan areas, creating, amongst others, the CBSA geography. OMB frequently refers to CBSAs as Metropolitan Statistical Areas (MSA), but since MSAs were defined differently prior to 2003, I use the CBSA moniker for clarity. For a thorough explanation of the changes, see the Missouri Census Data Center website (http://mcdc.missouri.edu/allabout/sumlevs/).

[28] I plan to expand the analysis to include additional metropolitan areas in future work. This will allow me to include measures of commuting and congestion that vary across metropolitan areas in estimation. Ideally, I would define the single market as the the Washington-Baltimore-Northern Virginia, DC-MD-VA-WV CSA, as the Washington, DC and Baltimore, MD residential and labor markets are undoubtedly linked. Doing so would drastically increase the number of Census blocks in the market and the scope of the GIS network analysis (that requires calculating the optimal route between all pairwise combinations of blocks in the market). It is not feasible at this time.

[29] I define the tails of the income distribution net of housing costs as below the 15th percentile or above the 90th percentile. The former restriction was chosen to ensure that all individuals have positive consumption while living in their observed homes and commuting by their observed modes.

that contain an unemployed, military, or part-time employed adult.[30] I also drop households that contain an adult whose job location information is missing or indicates that the individual works outside the geographic scope of the market. Next, I drop households with individuals who commute by methods that are either unavailable in the market (streetcar), occur too infrequently in the data to be modeled as outcomes (bike, commuter rail, ferry, taxi, motorcycle, other), or are beyond the scope of the model (working at home). Finally, I drop individuals who travel for an extremely long time or who cover an implausibly long distance as part of their commute.[31]

Based on the publicly available version of the ACS, the percent of the sample dropped for each specific reason is detailed in Table 3.[32] Regardless of whether the reason for the drop is a household or individual level characteristic, I drop the entire household. Column (1) contains the percent of households dropped for the given reason, and column (2) contains the analogous percent of individuals dropped.

7 Choice Set

After dropping individuals who commute by unavailable or infrequently observed methods, I model five commuting options in DC CBSA: automobile, carpool, Metrorail (heavy rail), Metrobus (bus), and walking. A key shortcoming of the ACS commuting data is that it reports only the primary method of travel one uses to commute, so I treat individuals in the model as if they do not commute by multiple modes.[33]

Table 4 shows the distribution how individuals in the data commute before and after sample selection.[34] Columns (1) and (2) are calculated from the full sample. Column (1) contains the percent of individuals who commute by the given method, and column (2) contains the standard deviation. Columns (3) and (4) contain the analogous figures for the selected sample. Households in the selected sample are about 20 percentage points more likely to commute by automobile, most likely because of the income selection criteria and the employment and commuting requirements that shift individuals out of the “other or not in labor force” category. The “other” component of that category represents the five unmodeled commute modes listed in Table 3.

I define the choice set in my discrete choice model as the $N^K = 5$ commuting options available

[30] The “head of household in school” reason in the summary table indicates households where the head of household is enrolled in grade school or a lower grade.

[31] Based on the distribution of commute times in the publicly available data, I define such a commute as one with a duration greater than the 95th percentile regardless of mode. This cutoff corresponds to a commute of approximately 75 minutes. Based on the distributions in the restricted-access data, I also restrict commutes that cover a distance greater than the 90th percentile for those who walk to work or a distance greater than the 90th (95th) percentile of the distance between the home (work) location and the nearest heavy rail station.

[32] These summary statistics are based on the 2005-2008 Public Use Microdata Sample (PUMS) version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns. As I do not observe individuals’ home and work locations at with precision in the PUMS data, I am unable to report how many observations are dropped due to these criteria.

[33] See Appendix A.1 for a description of this issue and what I do to mitigate the problem.

[34] These summary statistics are based on the PUMS version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns.

Table 3: Percent of Sample Dropped by Reason

Variable	(1) Percent of Households	(2) Percent of Individuals
Housing Unit Characteristics (H_{ih})		
Vacant house	0.045	0.000
Mobile home or RV	0.012	0.010
No cash rent	0.011	0.010
Meals included in rent	0.006	0.003
Household Characteristics (X_i)		
Net of housing exp. income tails	0.257	0.209
Subfamilies in household	0.111	0.111
Roomate present	0.026	0.015
Under 18 non-children	0.001	0.000
Adult children	0.118	0.060
Child primary wage earner	0.019	0.008
Individual Characteristics (X_i)		
Unemployed or not at work	0.226	0.130
Military employment	0.018	0.008
Not full time and year emp.	0.310	0.149
Head of household in school	0.001	0.000
Job location missing	0.030	0.014
Job location outside market	0.617	0.385
Commute Modes (k)		
Commute by streetcar	0.001	0.000
Commute by bike	0.005	0.002
Commute by commuter rail	0.008	0.004
Commute by other method	0.011	0.005
Work at home	0.052	0.024
Commute Characteristics (K_{ihk})		
Commute duration in tail	0.008	0.003
HR station (home) distance in tail	-	-
HR station (job) distance in tail	-	-
Walk distance in tail	-	-

Table 4: Percent of Commuters by Mode (k)

	(1)	(2)	(3)	(4)
	All Observations		Selected Sample	
Variable	Percent	SD	Percent	SD
Automobile	0.486	0.500	0.686	0.464
Carpool	0.072	0.258	0.078	0.267
Heavy Rail (Metrorail)	0.053	0.223	0.035	0.183
Bus (Metrobus)	0.016	0.126	0.020	0.140
Walking	0.016	0.126	0.020	0.140
Other/Not in LF	0.349	0.477	0.164	0.370

Notes: These summary statistics are based on the PUMS version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns.

in the DC CBSA and the N^H homes observed in the data. This means that there are potentially $N^K \times N^H$ options in an individual's choice set.[35] As Vega and Reynolds-Feighan (2009) explain, the econometrician needs to limit the size of the choice set when dealing with a large number of housing alternatives to make estimation tractable.[36] I address this issue in my model by randomly sampling to reduce the the number of households included in the sample (and thus individuals and housing options as well). Specifically, I randomly select $10,000$ households from the sample to form a new selected, random sample. Doing so allows me to avoid arbitrary spatial aggregation of housing alternatives that would reduce the precision with which I am able to map commuting alternatives.

8 Summary Statistics

Table 5 shows the distribution of housing characteristics before and after sample and random selection. Again, columns (1) and (2) are calculated from the full sample and columns (3) and (4) are based on the selected sample. The selected sample does not differ greatly from the full sample. The estimates describe what type of building the home is. The majority of homes in the sample are single family detached homes, as 58 percent of the households in the selected sample are of that type. Homes in the samples have an average of just over six rooms and are about 30 years in age.

Table 6 contains the analogous moments for key individual and household characteristics. The sample is slightly more female than male. Individuals in the selected sample average 45 years in age, have lived in their home for just under 9 years, have 0.7 children living in the home, and own 2 cars.

[35] I say, "potentially" because I allow individuals in the model to commute by automobile only if they own an automobile and limit housing options to homes that individuals can afford.

[36] Vega and Reynolds-Feighan explain two possible alternatives to dealing with this issue: restricting the individual's choice set to a random sample of all alternatives or spatially aggregating home location alternatives. The former requires restrictive assumptions on the error structure (see McFadden 1978) and the later is problematic because the unit of aggregation is arbitrarily defined. Since neither is tenable in my model, I propose a third alternative.

Table 5: Moments of Housing Characteristics (H_{ih})

	(1)	(2)	(3)	(4)
	All Observations		Selected Sample	
Variable	Percent	SD	Percent	SD
Single-Family Home-Detached	0.549	0.498	0.577	0.494
Single-Family Home-Attached	0.185	0.388	0.178	0.383
10+ Apartments	0.186	0.389	0.293	0.455
Property Age	33.474	19.440	29.035	2.036
Number of Rooms	6.328	2.167	6.046	2.036

Notes: These summary statistics are based on the PUMS version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns.

Table 6: Moments of Individual and Household Characteristics (X_i)

	(1)	(2)	(3)	(4)
	All Observations		Selected Sample	
Variable	Percent	SD	Percent	SD
Individual Characteristics				
Male	0.463	0.499	0.484	0.500
College Diploma +	0.511	0.500	0.434	0.496
Individual's Age	49.143	15.215	44.769	11.994
Household Characteristics				
Owner Occupied	0.781	0.414	0.780	0.414
Tenure in Home	10.011	9.464	8.636	8.405
Child in Home	0.356	0.479	0.383	0.486
Number of Children	0.649	1.022	0.709	1.053
Number of Vehicles	2.003	1.082	2.009	0.955

Notes: These summary statistics are based on the PUMS version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns.

9 Census Geography Background

Before explaining how I augment the ACS data with characteristics of the commute using GIS, I first provide detail on the Census Geography that forms the basis for the procedure. Census geography is complex because it deals with geographic entities that are both determined by legal boundaries that the Bureau does not control (counties, congressional districts, school districts, etc.) and Census defined summary areas (Census Blocks, PUMAs, etc.) that are used to report statistics at varying levels of aggregation. These geographic entities range in size from the Census block, which is the lowest level of Census geography, to the nation as a whole.[37] In ascending order of size, the geographic entities that are relevant for my analysis are: Census blocks, block groups, tracts, and Public Use Microdata Areas (PUMAs).

Census blocks are defined to serve as the building blocks of all other Census geographies and all land in the United States is assigned to a Census block. They are bounded on all sides by physical features (such as roads or streams) or invisible boundaries (such as city or county limits). They are generally geographically small, but can be large in unpopulated areas. Census blocks are clustered into slightly larger block groups, which in turn are clustered into Census tracts. Tracts are created to contain 4,000 individuals, although they range in size from 1,500 to 8,000 people nationally. They are defined to provide a consistent geographic unit for the Census to use to present aggregate statistics. Finally, PUMAs are areas defined to contain at least 100,000 people and are so created to ensure confidentiality in individual level data.

Table 7 contains the number of these geographies that fall inside the boundaries of the DC CBSA and their mean size.[38] Figure 1 presents the same information visually.

Table 7: Geographies in the DC CBSA

Variable	(1) Count Sum	(2) Size Mean
DC CBSA	1	6,030.347
States	4	1,507.587
Census PUMAs	45	134.008
Census Tracts	1,040	5.798
Census Block Groups	2,979	2.024
Census Blocks	51,972	0.116

Source: Author's calculations.
Note: The unit of measurement for size is square miles.

[37] I again refer the interested reader to the Missouri Census Data Center website for more detail on this topic (http://mcdc.missouri.edu/allabout/sumlevs/).

[38] I also include states in the table for reference.

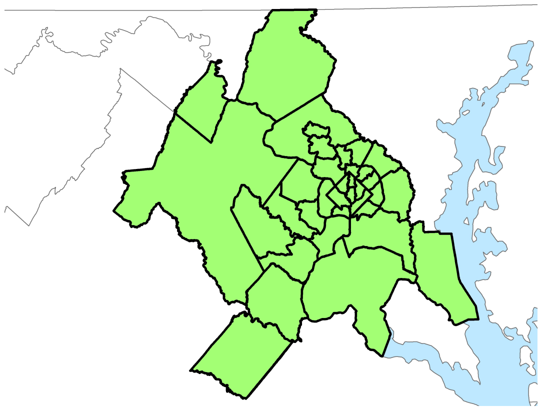

Figure 1a: Census PUMAS

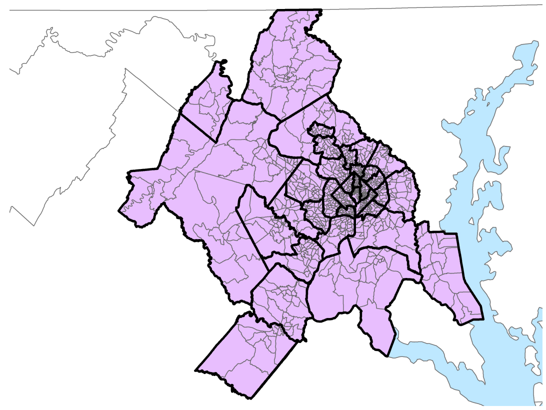

Figure 1b: Census Tracts

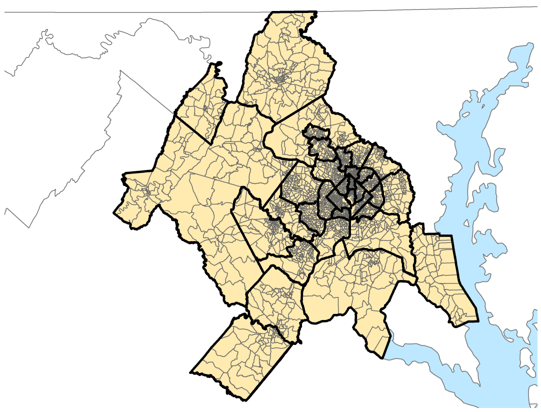

Figure 1c: Census Block Groups

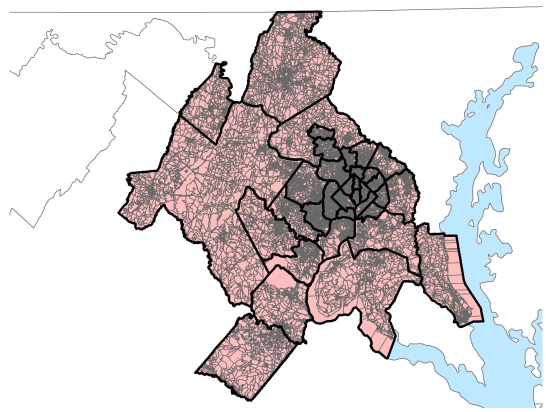

Figure 1d: Census Blocks

Figure 1: Geographies in the DC CBSA

Individual data is not publicly available at the tract level or below. The smallest geographic identifier in publicly available Census microdata is the PUMA, which averages 134 square miles in the DC CBSA. The restricted-access version of the ACS contains geographic information down to the block level. At an average of 0.12 square miles in size in the DC CBSA, Census blocks allow for much greater geographic precision in mapping the locations of individual residential and job locations. This precision is particularly important when mapping locations relative to the commuting infrastructure (such as highways or Metro stations) in the market. Attempting to map the commute between areas that are 134 square miles in size would be an imprecise exercise at best and an impossible exercise for individuals who live and work in the same PUMA. Thus, the geographic precision available in the restricted-access ACS data is essential for creating the GIS data that as accurately as possible approximates the characteristics of both observed and unobserved commutes.[39]

[39]In order to replicate the conditions in the RDC based on the PUMS data for preliminary analysis, I randomly assign households to a population weighted residential block location within their reported PUMA. The residential population weights are based on available block level aggregate population counts. I also randomly assign individuals a job block location within their reported PUMA, but analogous employment density weights are not readily available.

10 GIS Data Calculation

To perform the GIS network analysis that allows me to calculate the optimal route between a home and job location pair, I begin by constructing a digital representation of the Census geography.[40] I use the 2009 definition of the CBSA from the Census TIGER/Line® shapefiles to define the market. Since blocks, block groups, and tract definitions are updated every Decennial Census, I use ESRI ArcGIS software to keep all of the 2000 definition of the Census TIGER/Line® block, block group, and tract shapefiles that fall within the boundaries of the DC CBSA. I use block centroids to approximate the exact home or job location. Next, I overlay a street network and a rail network on the Census geographies.[41] The street network data is obtained from ESRI's Data & Maps 9.3 (StreetMap North America). The rail network is created from the locations of rail stations and lines available in the National Transportation Atlas Database (NTAD) for both heavy rail (Metrorail) and commuter rail (MTA and VRE). Both sources are updated infrequently, so I use one version of the network as the basis for the analysis, as opposed to creating multiple, year specific networks.[42]

For each job location in the CBSA, I calculate the optimal route from that job location to every home location by every commuting method.[43] I do not observe transfers in the data, so I only need commutes by each given method, not the optimal combination of the methods. Optimal routes are calculated using the ArcGIS OD Cost Matrix Solver, which uses a version of Dijkstra's Shortest Path Tree algorithm to search for the lowest time cost route on a network between two points.[44] The optimization takes into account turns, stops, and speed limits for automobile travel and stops, transfers at defined hubs, and average speeds for rail travel.[45]

For the calculated optimal route by road travel methods between the home and job locations,

[40] All GIS boundary files (shapefiles) were downloaded from the National Historical Geographic Information System (NHGIS) website (www.nhgis.org).

[41] I have not yet created the bus network, but plan to do so using bus station and line information from DC GIS. Currently, I use calculations from the automotive network as a proxy for bus commute characteristics.

[42] The street network is based on 2003 TeleAtlas data. Heavy rail (subway) information comes from the 2004 Fixed-Guideway Transit Network database created by the University of Tennessee Center for Transportation Research GIS Group. The commuter rail network is created from data compiled by the Research and Innovative Technology Administration's Bureau of Transportation Statistics (RITA/BTS) for the 2009 NTAD. All sources used were the most current data available at the time of the construction of the network.

[43] Since I do not observe actual home and job locations outside the RDC, and GIS capabilities are limited in the RDC, I have to calculate the routes between all pairwise combinations of locations outside the RDC and import the resulting data. Doing so using Census blocks would require calculating $\frac{51{,}972\times(51{,}972+1)}{2} \approx 1.35\, billion$ automotive routes. This is beyond the GUI capabilities of ArcGIS, but can be accomplished by writing a Python script that accesses the GIS processor and loops over locations. To reduce the dimension of the computational burden and the size of the data, I take advantage of the fact that some block groups and tracts are very large in geographic size relative to their component blocks, while other block groups and tracts are not much larger than their component blocks. I develop algorithm that selects the largest Census geography (block/block group/tract) that will give a reasonably precise measure of location in order to balance computational burden and data size against precision.

[44] The algorithm simultaneously solves forward from the origin and backwards from the destination (in a hierarchical fashion for roads) until the two paths meet. See Houde (2012) for technical details of how the algorithm works.

[45] Speed limit information is contained in the street network data. Average rail speeds are approximated based on the author's calculations from Metrorail, MTA, and VRE schedules.

I am able to calculate the distance traveled on the network and the predicted travel time if one travels the speed limit. These distances and times should be thought of as similar to the ones an individual would recover from an online mapping website or a GPS, so it is important to note that they do not account for congestion.[46] For rail travel methods, I calculate the analogous distance and the travel time if one travels the average speed. As discussed in Appendix A.1, to control for the fact that the ACS only reports the primary method of travel, I also calculate the as-the-crow-flies distance from both home and job locations to the nearest rail station. Finally, for individuals who commute by walking, I also calculate the as-the-crow-flies distance between locations to provide information about the characteristics of their commute, as there is no geographic network applicable to pedestrians.

Table 8 shows the distribution of key ACS and GIS commuting characteristics before and after sample selection. All commute times are in hours per week. The average commute time

Table 8: Moments of Commute Characteristics

	(1)	(2)	(3)	(4)
	All Observations		Selected Sample	
Variable	Mean	SD	Mean	SD
ACS Commute Characteristics (t_{ihk})				
Commute time	5.369	4.109	4.890	3.773
Auto time	5.239	3.741	4.622	3.631
Carpool time	6.279	4.053	6.111	3.642
Metrorail time	7.607	3.453	7.684	3.534
Metrobus time	8.082	4.848	8.194	5.357
Walk time	2.259	2.128	1.611	1.666
GIS Commute Characteristics (K_{ihk})				
GIS auto commute time	-	-	-	-
GIS Metrorail commute time	-	-	-	-
Metrorail station to home (mi)	-	-	-	-
Metrorail station to job (mi)	-	-	-	-
ATCF distance (mi/week)	-	-	-	-

Notes: The unit of measurement for time is hours/week. The ACS commute characteristic statistics are based on the PUMS version of the ACS (as opposed to the restricted-access version) to mitigate disclosure risk concerns. The GIS commute characteristic statistics require restricted-access data to be meaningful and have not yet passed disclosure review.

reported in the ACS random sample is 4.9 hours per week. This is similar to the average automotive

[46]Although possible, I do not repeatedly query an online mapping website and record the resulting data. Small scale experiments with such a process using Google Maps were slower than using GIS network analysis, and a mass download of the amount of data I would need would require prior approval from Google to avoid violating their Terms of Service.

commuting time of 4.6 hours per week. Commuting by carpool results in a longer commute of 6.1 hours per week on average, as would be expected. Commuting by mass transit results in an average commute of around 8 hours per week, while walkers have the shortest commutes, on average, likely due to the fact that only those with short distances to travel can plausibly walk to work. The average commute times and distances predicted by the GIS network analysis have not yet been cleared for disclosure.

Table 9 presents log-linear Ordinary Least Squares (OLS) regressions of reported ACS commute times on predicted GIS commute characteristics, conditional on traveling by the given commuting method. The independent variables are the GIS network travel time for the given mode,

Table 9: Baseline Log-Linear Commute Time Regressions

	(1)	(2)	(3)	(4)	(5)
	Auto	Carpool	Metrorail	Metrobus	Walk
Variable	$\ln(t_{ihk})$	$\ln(t_{ihk})$	$\ln(t_{ihk})$	$\ln(t_{ihk})$	$\ln(t_{ihk})$
GIS Measure	0.145***	0.095***	0.161***	0.073***	0.164***
	(0.002)	(0.005)	(0.005)	(0.007)	(0.009)
GIS Measure Squared					-0.005***
					(0.000)
Metrorail Station to Home (mi)			0.254***		
			(0.020)		
Metrorail Station to Home Squared			-0.049***		
			(0.005)		
Metrorail Station to Job (mi)			0.418***		
			(0.100)		
Metrorail Station to Job Squared			-0.328***		
			(0.122)		
Constant	0.851***	1.211***	1.193***	1.566***	-0.027
	(0.008)	(0.029)	(0.019)	(0.021)	(0.037)
N (Rounded)	14,000	800	3,100	1,200	1,100
R^2	0.334	0.305	0.426	0.080	0.350

Notes: The unit of measurement for time is hours/week. The GIS measure is commute time for all categories, save walking, where it is the as-the-crow-flies distance (mi/week). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

save for walking. I use the as-the-crow-flies distance between the home and work locations to inform pedestrian commute times. These simple regressions show that the calculated commute times are all positive, significant predictors of the commute times reported in the ACS. Unfortunately, there is a great deal of the variation in commute time around the mean that I am not explaining, as evidenced by the low values of the R^2s. This is likely the result of two shortcomings of the estimation. I have already mentioned the first: unreported multimodal commuting methods. More importantly, I have not yet developed an appropriate measure of congestion to include in the model. There is a great deal of congestion in Washington, DC, so this is likely to affect the fit of the model. That the coefficient on Metrorail is much closer to one than the other coefficients supports this hypothesis, as congestion is much less likely to cause delays on subways that run on fixed schedules. Regardless of the deficiency, these regressions show that the GIS network analysis does a reasonable job of modeling commute characteristics.

11 Pricing Data

I also augment that dataset with pricing information from multiple data sources. Olsen et al. (2012) provides measures of the price of composite consumption (p_m^c) in the form of a price index for non-housing goods in the given year. Although I do not have variation in markets that would necessitate the use of this index, my data is a repeated cross section, so I include this measure to smooth variation in prices over time.

I construct the opportunity cost of living in each home (p^H) by modifying a procedure outlined in Bayer et al. (2005) and Bayer et al. (2007). The details of this procedure can be found in Appendix A.2.

Finally, for the per mile price of each commuting method (p_{mk}^d), I use data from the Energy Information Administration (EIA) for gas prices and the National Transportation Database (NTD) for average fares. Gas prices based on the average annual regular reformulated retail gas price in dollars per gallon for the lower Atlantic region. I calculate average fares from the NTD by dividing total annual fares collected by total annual passenger miles for the given mode.

Part V

Estimation

I develop an empirical specification that modifies McFadden's (1974) RUM model so that the assumption of a linearly additive error term in the utility function is not required and allows for joint estimation of discrete outcomes (housing location and commute mode) and continuous outcomes (commute times) based on those decisions. My approach makes use of many of the discrete choice tools described in Train (2009), and I estimate my model with the Maximum Simulated Likelihood (MSL) methods of Geweke (1989), Hajivassiliou (1990), Keane (1994) (GHK), and Stern (1997). The individual likelihood contribution is the probability of observing the sample data given the parameters (θ) of the model. Simulation is required to evaluate these probabilities because they contain multidimensional integrals over the joint distribution of the errors that cannot be evaluated analytically.

This section proceeds by explaining the empirical specification. I then discuss how the parameters in the model are identified and the potential biases introduced by the assumptions I make.

12 Empirical Specification

The likelihood contribution involves three dependent variables. Define the observed housing choice of individual i as h and her observed commuting method as k. Let $P_i = \Pr(h, k, t_{ihk} \mid \theta)$ denote the probability of observing individual i living in home h and commuting by method k for a duration of t_{ihk} conditional on the parameters in the model. I proceed as follows: first, I define the structure of the errors in my theoretical model. Next, I show that P_i can be decomposed into the product of the joint probability of observing an individual living in house h and commuting by method k and the probability of commuting for a duration of t_{ihk}. I then detail the estimation routine for each factor separately. Finally, after explicitly defining each of those terms, I am able to write the likelihood function and its simulated analog. Note that, in this section, I drop the m subscripts for notational convenience since I am currently only using the DC CBSA in estimation.

12.1 Error Structure

There are two types of errors in the single person family utility functions: an idiosyncratic error and unobserved heterogeneity terms. The former, e_{ihk}, accounts for the difference between the predicted and observed commute times. The latter comes in three forms: 1) μ_i and u_i are the unobserved components of preferences for housing amenities and leisure time, 2) ε_{ih} is the unobserved component of the value of house h, and 3) λ_{ik} is the unobserved time value of commuting method k. These errors are assumed to be known to the agents but not the econometrician. I proceed by first discussing the idiosyncratic error term, then discussing the role the unobserved heterogeneity terms play in estimation.

12.1.1 Idiosyncratic Error

Since t_{ihk} is not a choice variable, but rather is determined by the choices of h and k, then any deviation in the predicted t_{ihk} from the true travel time is assumed to be idiosyncratic. I assume this error is distributed as $e_{ihk} \sim iidN(0, \sigma_e^2)$.[47] This adds a variance parameter to the model, σ_e^2.

12.1.2 Unobserved Heterogeneity

Define each of the three types of unobserved heterogeneity errors and their distributions as $\vec{\mu}_i = (\mu_i, u_i) \sim N\left(0, \Omega^{\vec{\mu}}\right)$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iN^H}) \sim N(0, \Omega^{\varepsilon})$, and $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{iN^K}) \sim LN\left(0, \Omega^{\lambda}\right)$. The latter distribution is chosen to ensure that λ_{ik} is bounded below at 0, as is required by the theoretical model. In order to both normalize the model and reduce the computational burden of the estimation routine while still retaining a rich set of covariance terms, I impose structure on Ω^{ε} and Ω^{λ}. I do so by defining ε_i^H and λ_i as being functions of correlated and idiosyncratic components, in ways that still allow for substantial correlations across related choices. Specifically, I

[47] Although it simplifies simulation of the choice probabilities, the assumption that the e_{ihk} are iid is not necessary for estimation. A more complex correlation structure can be accounted for with a GHK simulator.

assume that unobserved preferences for homes are correlated for the same individual within and across neighborhoods, but not across homes themselves. Similarly, unobserved preferences for time spent commuting are correlated within commuting method classifications, but neither across classifications nor individual commuting methods. I explain these restrictions in greater detail in the subsequent paragraphs. I detail the specification of λ_i first, as it is more straightforward.

Let $\tilde{\lambda}_{i\tilde{k}}$ be an error associated with traveling by commuting method category $\tilde{k}$ and w_{ik} an idiosyncratic error associated with commute method k. Formally, assume that $\lambda_{ik} = \exp\left(\tilde{\lambda}_{i\tilde{k}} + w_{ik}\right)$, where $\tilde{\lambda}_{i\tilde{k}} \sim iidN\left(0, \sigma^2_{\tilde{\lambda}}\right)$ and $w_{ik} \sim iidN\left(0, \sigma^2_w\right)$.[48] I assume there are three commuting method categories: personal, mass transit, and other; with the "Car, Truck, or Van" and "Carpool" commuting methods belonging to the first category, the "Bus" and "Subway" commuting methods belonging to the second, and the "Walk" category belonging to the last. The intuition behind these classifications is best explained with an example. Individuals who have a high taste for the convenience and flexibility of driving one's own automobile to work (for instance, the ability to park near one's origin and destination) are also likely to have a high taste for the relative convenience and flexibility of carpooling (the ability to park or be picked up and dropped off near one's origin and destination). This would be evidenced in the model by the fact that the errors associated with "Car, Truck, or Van" and "Carpool" would be correlated through their common $\tilde{\lambda}_{i\tilde{k}}$ term. This specification sacrifices some flexibility, but still retains much of the important detail of the model and reduces the number of parameters in Ω^{λ} from 28 to 2.

Similar to $\tilde{k}$, let $\tilde{h}$ index neighborhoods and $\tilde{\varepsilon}_{i\tilde{h}}$ be the component of the error associated with neighborhood $\tilde{h}$. Assume that $\varepsilon_{ih} = \tilde{\varepsilon}_{i\tilde{h}} + \upsilon_{ih}$, where the first term is allowed to be correlated with other members of its group and the second term is idiosyncratic: $\tilde{\varepsilon}_i \sim N\left(0, \Omega^{\tilde{\varepsilon}}\right)$ and $\upsilon_{ih} \sim iidN\left(0, \sigma^2_{\upsilon}\right)$.[49] Previous studies have defined neighborhoods based on Census geography at either the Census Block, Block Group, or Census Tract level.[50] There are 51,972 Census Blocks, 2,979 Block Groups, and 1,040 Census Tracts in the DC CBSA. Estimation of the $\frac{N^{\tilde{H}}\left(N^{\tilde{H}}+1\right)}{2}$ elements in $\Omega^{\tilde{\varepsilon}}$ at even the Census Tract level is computationally infeasible. In order to allow for covariation in neighborhood unobservable characteristics in an estimable manner, I define the correlation between any two given neighborhoods as being a decaying function of the distance between those neighborhoods. The intuition behind this specification is that the unobservable characteristics of two neighborhoods that are one mile apart should be more closely correlated than the unobservable characteristics of two neighborhoods that are five miles apart, and beyond a threshold distance, there should be no correlation. This specification assumes that there is an underlying

[48] Note that this preserves the log-normal distribution of λ_i because the sum of two normally distributed random variables is normally distributed, and the exponent of a normally distributed random variable is log-normally distributed.

[49] Since N^H is large, assuming that $\upsilon_{ih} \sim iidN\left(0, \Omega^{\upsilon}\right)$ would be intractable because it would mean that there are N^H variance parameters to estimate.

[50] Bayer et al. (2008) find evidence of neighborhood effects in hiring networks at the Census block level. Bayer et al. (2004) uses Census block groups to define neighborhoods when examining racial segregation in housing markets. Bayer et al. (2007) uses school attendance zones, as well as including controls at both the Census block and block group levels. Kiel and Zabel (2008) find that multiple definitions of a neighborhood, including Census tracts, are jointly relevant in hedonic equations.

continuum of unobservable neighborhood characteristics that dies out as distance from the given neighborhood increases, as opposed to a discrete change in unobservable characteristics when one crosses from the given neighborhood to "the other side of the tracks." I define neighborhoods at the Census Tract level and let $\tilde{h}$ and $\tilde{j}$ index neighborhoods. I also define $\tilde{d}_{\tilde{h}\tilde{j}}$ to be the "as-the-crow-flies" distance between the given neighborhoods. I define a spline function that weights the correlation between the $\tilde{h}$th and $\tilde{j}$th neighborhoods as

$$\begin{aligned} \delta_{\tilde{h}\tilde{j}}\left(\tilde{d}_{\tilde{h}\tilde{j}}\right) &= 1\left(\tilde{d}_{\tilde{h}\tilde{j}}=0\right)\delta_0 + 1\left(0<\tilde{d}_{\tilde{h}\tilde{j}}\leq 1\right)\tilde{d}_{\tilde{h}\tilde{j}}\delta_1 \\ &+ 1\left(1<\tilde{d}_{\tilde{h}\tilde{j}}\leq 3\right)\tilde{d}_{\tilde{h}\tilde{j}}\delta_2 + 1\left(3<\tilde{d}_{\tilde{h}\tilde{j}}\leq 5\right)\tilde{d}_{\tilde{h}\tilde{j}}\delta_3, \end{aligned}$$

where $1(\cdot)$ is an indicator function equal to 1 if the argument is true and 0 otherwise. Note that $\tilde{d}_{\tilde{h}\tilde{j}}=0$ implies that $\tilde{h}=\tilde{j}$. Since manipulating a 1040×1040 matrix is computationally costly, I do not estimate $\Omega^{\tilde{\varepsilon}}$. Instead, I allow neighborhood unobservables to be correlated by calculating $\tilde{\varepsilon}_{i\tilde{h}}$ as a weighted sum of the standard normal errors associated with each neighborhood. By defining $\eta^{\tilde{\varepsilon}}_{i\tilde{j}}\sim iidN(0,1)$, the definition of $\tilde{\varepsilon}_{i\tilde{h}}$ can be stated formally as $\tilde{\varepsilon}_{i\tilde{h}}=\sum_{\tilde{j}=1}^{1040}\delta^{\tilde{\varepsilon}}_{\tilde{h}\tilde{j}}\eta^{\tilde{\varepsilon}}_{i\tilde{j}}$. This specification reduces the number of variance/covariance parameters in $\Omega^{\tilde{\varepsilon}}$ to be estimated to four (the elements of the vector δ).

After specifying the errors in this way, there are five vectors of unobserved heterogeneity errors. To keep subsequent notation compact, I define a vector of the unobserved heterogeneity terms as $\xi_i=\left(\vec{\mu}_i,\tilde{\varepsilon}_i,\upsilon_i,\tilde{\lambda}_i,w_i\right)$. Let $\theta=\left\{\beta^H,\beta^\ell,\gamma^H,\gamma^K,\sigma_e,\Omega^{\vec{\mu}},\delta,\sigma_\upsilon,\sigma_{\tilde{\lambda}},\sigma_w\right\}$ be the full set of parameters to be estimated. After imposing structure on the errors in my model, I am able to reduce the total number of variance/covariance parameters to 11.

12.2 Joint Probability of Observing t_{ihk}, h, and k

Next, I use the error structure to define the probability of observing the sample data, $P_i=\Pr(h,k,t_{ihk}\mid\theta)$. Using the law of total probability, I write this joint probability as

$$P_i = \Pr(t_{ihk}\mid\theta)\Pr(h,k\mid e_{ihk},\theta), \tag{3}$$

since e_{ihk} is the only random component of t_{ihk}. The first factor is the probability of observing individual i commuting for a duration of t_{ihk}. Explicitly,

$$\Pr(t_{ihk}\mid\theta)=\Pr\left(t_{ihk}=\exp\left(K_{ihk}\gamma^K_k+e_{ihk}\right)\ for\ h\ and\ k\mid\theta\right).$$

The second factor is the probability of observing individual i living in house h and commuting by method k, so $\Pr(h,k\mid e_{ihk},\theta)=\Pr(U_{ihk}>U_{ih'k'}\,\forall\,(h',k')\neq(h,k)\mid e_{ihk},\theta)$. For notational compactness, I define $\Pr(t_{ihk}\mid\theta)\equiv P^t_i$ and $\Pr(h,k\mid e_{ihk},\theta)\equiv P^{HK}_i$. Appendix B.1 details how P^t_i and P^{HK}_i are calculated.

12.3 Likelihood Function

Assume that there are N total individuals in the data. The log likelihood function is

$$\ln L(\theta) = \sum_{i=1}^{N} \ln\left(P_i^t P_i^{HK}\right). \tag{4}$$

12.4 Simulation

Evaluation of the multidimensional integrals in $L(\theta)$ is not possible analytically or numerically, so I use a GHK simulator to evaluate the choice probabilities. In Appendix B.1, I define B_i as the upper bound on ξ_i, $\bar{\phi}(\xi_i)$ as the truncated, joint distribution of ξ_i, and $P_i^B = \Pr(\xi_i < B_i)$ as the probability that ξ_i is less than B_i. Following Stern (1997), I compute N^R draws of $\xi_{ir} \in B_i$ from $\bar{\phi}(\xi_i)$. I define $P_{ir}^{BR} = \Pr(\xi_{ir} < B_{ir})$ as the simulated analog to P_i^B (Appendix Equation 6).[51] I replace the analytical likelihood contribution of P_i^{HK} (Appendix Equation 7) with its unbiased simulated analog as

$$\begin{aligned} P_i^{HKR} &= \frac{1}{N^R} \sum_{r=1}^{N^R} \prod_{(h',k') \neq (h,k)} \left[1 - \Phi\left(\frac{f(\xi_{ir}, D_{ih'k'})}{\sigma_e}\right)\right] P_{ir}^{BR} \\ &= \frac{1}{N^R} \sum_{r=1}^{N^R} P_{ir}^{HKR}, \end{aligned}$$

where $f(\xi_{ir}, D_{ih'k'})$ is shorthand defined in Appendix B.1.2 such that $e_{ih'k'r} > f(\xi_{ir}, D_{ih'k'})\ \forall\ (h',k') \neq (h,k) \Leftrightarrow U_{ihkr} > U_{ih'k'r}\ \forall\ (h',k') \neq (h,k)$.

The simulated likelihood function is

$$\ln L(\theta) = \sum_{i=1}^{N} \ln\left(\left[\frac{1}{N^R} \sum_{r=1}^{N^R} P_{ir}^{HKR}\right] P_i^t\right), \tag{5}$$

and estimation proceeds by MSL. I maximize the simulated likelihood function using the optimization routine outlined in Berndt, Hall, Hall, and Hausman (1974) which is commonly referred to as the BHHH algorithm.

[51] More precisely, the $\Pr(\xi_i < B_i) = \Pr(w_{ik} < B_i^w)$ because only one element of ξ_i is bounded, so $P_{ir}^{BR} = \Pr\left(w_{ikr} < B_i^w\left(\tilde{\lambda}_{i\tilde{k}r}\right)\right)$, but these details are only relevant to readers of Appendix B.1.

13 Identification

Having explained how I estimate my joint model of residential choice and commuting mode, I proceed by providing details on how the parameters in the model are identified. In this section, I address the need for a structural model and how I separately identify the effects of the endogenous decisions in the model. Appendix B.2 details the variation that identifies each parameter in the model, and Appendix B.3 discusses threats to identificaiton.

13.1 Why a structural model?

In thinking about identification, it is important to first explain the need for the econometric sophistication used in my model. Quite simply, this is due to the fact that a randomized, controlled experiment that would address my research question would be impossible to implement, and no natural experiment exists that would allow me to disentangle the separate effects the many relevant factors and motivations have on observed responses.[52] Being able to explain why individuals react the ways they do is important for predicting responses to the policies I am investigating. For instance, Baum-Snow and Kahn (2000) use panel data methods to examine how much expansions to rail transit systems cause individuals to switch to commuting by rail. They find evidence of an increase in rail transit use in areas near expansions, but they cannot determine what percentage of that increase is due to new riders and what percentage can be attributed to former rail commuters who moved from another location to take advantage of the infrastructure improvements.

While my structural model comes at a cost in terms of both implementation and understanding, it also yields important benefits. Estimation of preference parameters allows me to perform simulations that address a myriad of questions about the effects of proposed policies that have not yet been widely implemented. The model also allows for the extension of the collective model to a new arena in order to account for the fact that spouses behave differently than single individuals when making housing and commuting decisions. Addressing this issue directly would not be possible without a structural model.

13.2 Exclusion Restrictions

The main concern with a model of the joint decision of where to live and how to commute is that both decisions are made simultaneously. The housing location decision pins down where an individual is commuting from. Conversely, the availability and characteristics of commuting options are characteristics of the home themselves. In order to separately identify each effect without relying on functional form assumptions, I need at least one variable that exogenously affects each given decision alone. I use intrinsic, physical characteristics of the home that are observed in the data (e.g. number of rooms and property age) as an exclusion restriction to help identify the parameters

[52]To do so, a natural experiment would have to impact individuals in such a way that their responses would be through one of the channels I am modeling, but no others. With such interconnected decisions as residential location and commuting method, both of which are influenced by a multitude of factors, such an experiment is hard to imagine.

pertaining to the commuting mode decision.[53] I do not allow an individual's commute to factor into the decision to purchase a home beyond its effect on leisure, so I exclude commute characteristics from the housing equation to identify the parameters relevant to the decision to purchase a home.

Part VI

Results

This section summarizes the results from my estimated model. I proceed by first presenting the parameter estimates and standard errors. Since my model contains discrete outcomes, the parameter estimates cannot be interpreted as the effect of the explanatory variable on the outcome, so I also present accompanying marginal effects. The second section compares aggregate moments generated from the model with the true moments found in the data, and the third section more formally tests the model using several different specification tests. Finally, I discuss and perform policy simulations.

Although I have been able to estimate my model, I have not yet completed the calculation of all of the aforementioned results. Where necessary, I outline what will be calculated and presented as part of the published version of this research project.

14 Model Parameter Estimates

I present preliminary estimates from the model. I am in the process of calculating computationally tractable marginal effects, so these coefficients are estimates of utility parameters. They can be interpreted as affecting utility, but not the probability of choosing a particular home or commuting option. Although imperfect, these estimates are interesting, at face value because they indicate how observables affect utility. They are also useful as they provide evidence that the code that executes the estimation routine functions properly.

14.1 Leisure Parameter Estimates

Recall from Equation 2 and Section 12.1 that the commute time parameters are γ^K and σ_e, and the commute mode preference parameters are $\sigma_{\bar{\lambda}}$ and σ_w. Estimates of these parameters appear in Table 10. The first row contains parameters from the commute time equation. For each commuting method (save automotive commuting), I include a mode specific constant and the GIS predicted commute measure. I normalize the automobile constant to 0 for identification. As in Table 9, the GIS measure is predicted commute time for all measures, save walking. I use the as-the-crow-flies

[53] This is a valid restriction so long as the characteristics of homes vary with location. In other words, this restriction fails if one can buy an identical home in every location.

Table 10: Leisure Parameter Estimates

Variable	(1) Auto $\ln(t_{ihk})$	(2) Carpool $\ln(t_{ihk})$	(3) Metrorail $\ln(t_{ihk})$	(4) Metrobus $\ln(t_{ihk})$	(5) Walk $\ln(t_{ihk})$
Commute Characteristic Parameters (γ^K)					
GIS Measure	0.3710***	0.1440***	0.1030***	0.2660***	0.7680***
	(0.0017)	(0.0047)	(0.0179)	(0.0045)	(0.0408)
GIS Measure Squared					0.6010***
					(0.1140)
Metrorail Station to Home (mi)			0.3870***		
			(0.0660)		
Metrorail Station to Home Squared			-2.0280		
			(1.5020)		
Metrorail Station to Job (mi)			0.5890*		
			(0.3430)		
Metrorail Station to Job Squared			51.8300		
			(35.3200)		
Constant	0.000	2.783***	1.586***	2.412***	18.020***
		(0.0323)	(0.0725)	(0.0264)	(5.3130)
N (Rounded)	10,000				
Train's Pseudo R^2	0.25				

Notes: The unit of measurement for time is hours/week. The GIS measure is commute time for all categories, save walking, where it is the as-the-crow-flies distance (mi/week). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10. Cells without standard errors represent parameters that were normalized for identification.

Variable	(1) Estimate	(2) Std. Error	(3) t-Statistic	(4) p-Value
Second Moment Parameters (σ_e and Ω^λ)				
Std. Dev. of Commute Error (σ_e)	0.8240***	0.0000	22960.0000	0.0000
Std. Dev. of Mode Category Pref. Error ($\sigma_{\tilde{\lambda}}$)	0.4040***	0.0004	939.7000	0.0000
Std. Dev. of Mode Pref. Error (σ_w)	0.2540***	0.0006	454.1000	0.0000
N (Rounded)	10,000			
Train's Pseudo R^2	0.25			

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

distance between an individual's home and job to inform the walking commute time. I control for multimodal commuters who report commuting by subway by including a quadratic function of the distance to the nearest Metro station from both the home and work location in the Metrorail commute time equation. Finally,

As would be expected, increasing the GIS predicted commute time increases reported automobile commute times. The estimate of 0.371 is greater in magnitude to the estimate of 0.226 from the baseline commute time regressions (see Table 9), although the latter specification contained a constant. The analogous coefficients on carpooling and Metrorail are also positive, and similar in magnitude to the baseline specification. The distance to the Metrorail station effects indicate that time is a concave function of distance with a maximum of 0.095 (0.005) miles from home (job). This is consistent with individuals who need to travel greater distances to catch the subway traveling to the station by faster methods (ie, driving instead of walking). The Metrobus estimate is positive and significant, but almost four times greater than the estimate from the regression. This provides suggestive evidence that there is substantial endogeneity bias in the regression estimates among commuters by bus due to residential sorting. The as-the-crow-flies distance is a positive predictor of walk times, and the convex relationship between the two indicates that commute time increases at an increasing rate with distance walked. Finally, the standard deviation of the idiosyncratic commute time error, e, is the amount of variation in commute times that is not being explained by the model.

The standard deviation w is significant, indicating that there is substantial variation in how individuals view time spent commuting by different methods, but the significant standard deviation of $\tilde{\lambda}$ indicates positive correlations between the mode categories.

14.2 Housing Consumption Parameter Estimates

The housing parameters are γ^H, δ, and σ_υ. Estimates of these parameters and the associated standard errors are included in Table 11.

The first three parameter estimates describe the type of building the individual lives in. The baseline, omitted category is an apartment building with less than ten units. Unsurprisingly, single-family-detached homes are preferred to single-family-attached homes, given the relative magnitudes of the coefficients. Attached homes are are preferred to apartments of all sizes, as evidenced by the positive, significant coefficient estimates of the first two parameters and the insignificant coefficients for the third parameter. These are intuitive estimates, as are the results that that older structures decrease the utility one gets from the home and that living in a home with more rooms increases utility.

There are very small, insignificant effects for the δ parameters that govern how correlated the unobserved values of homes in neighborhoods in close proximity to one another are. This may be indicative that my specification of a neighborhood is incorrect or these amenities are imperfectly capitalized into the value of a home so the model cannot accurately determine the trade-offs that

Table 11: Housing Consumption Parameter Estimates

Variable	(1) Estimate	(2) Std. Error	(3) t-Statistic	(4) p-Value
Housing Characteristic Parameters (γ^H)				
Single-Family Home-Detached	0.1040***	0.0263	3.9650	0.0001
Single-Family Home-Attached	0.0511**	0.0228	2.2380	0.0253
10+ Apartments	0.0026	0.0189	0.1350	0.8920
Property Age	-0.1280***	0.0461	-2.7930	0.0052
Property Age Topcoded	0.0152	0.0231	0.6570	0.5110
Number of Rooms	0.0124***	0.0046	2.7060	0.0068
1-10 Acre Lot	0.1190***	0.0382	3.1230	0.0018
10+ Acre Lot	-0.0145	0.0985	-0.1470	0.8830
Second Moment Parameters (Ω^ε)				
Same Neighborhood (δ_0)	-0.0054	0.0065	-0.8370	0.4020
Neighborhoods within 1 Mile (δ_1)	-0.0002	0.0042	-0.0444	0.9640
Std. Dev. of Housing Amenity Error (σ_υ)	3.2780***	0.0253	129.6000	0.0000
N (Rounded)	10,000			
Train's Pseudo R^2	0.25			

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

individuals are making. The standard deviation of the idiosyncratic housing error is large relative to the parameter estimates and indicates the a large amount of variation in housing characteristics is not explained by the observed characteristics. I am in the process of adding more housing and neighborhood measures to address this issue.

14.3 Taste for Housing Consumption / Leisure Parameter Estimates

The preference parameters are those included in the αs: β^H, β^ℓ and $\Omega^{\vec{\mu}}$. Recall that the taste parameter that governs the relative weight the individual places on composite consumption (α_i^c) is normalized to one for identification. I present the parameters that govern relative taste for both housing and leisure in Table 12.

The positive, significant leisure constant indicates that leisure dominates consumption of both housing and all-other-goods. The lack of significance on the male coefficients indicate that men's and women's preferences do not differ substantially along these dimensions. All else equal, black individuals place more weight on composite consumption than housing and leisure relative other races, but there are no significant differences in preferences between Hispanics and non-Hispanics. Having a child in the home increases the value an individual places on both housing amenities and leisure time, consistent with a desire to provide a home for one's children and the time investment parents make in raising them. Individuals with college diplomas also place more weight on housing and leisure. Finally, as individuals age, they place less value on housing amenities and leisure

Table 12: Preference Parameter Estimates

	(1)	(2)	(3)	(4)
	Housing (α^H)		Leisure (α^ℓ)	
Variable	Estimate	Std. Error	Estimate	Std. Error
Demographic Characteristic Parameters	β^H		β^ℓ	
Male	-0.0108	0.0107	-0.0108	0.1240
Black	-0.8590***	0.0968	-1.2960***	0.1250
Hispanic	0.0479	0.3780	-0.1700	0.4340
Child in Home	0.7590***	0.1320	0.5690***	0.1540
College Diploma +	0.8450***	0.1120	0.7200***	0.1280
ln(Age)	-0.1060***	0.0295	-0.4330***	0.0986
Constant	0.0000		3.726***	0.3710
Second Moment Parameters ($\Omega^{\vec{\mu}}$)	μ		u	
Std. Dev. of Preference Errors (σ_μ and σ_u)	0.2060***	0.0468	0.2350***	0.0560
Correlation	0.9820***	0.0264		
N (Rounded)	10,000			
Train's Pseudo R^2	0.25			

Notes: *** p<0.01, ** p<0.05, * p<0.10. Cells without standard errors represent parameters that were normalized for identification.

(and more on composite consumption). These final estimates are not intuitive and require further examination.

The standard deviation of μ_i (u_i) is 0.2060 (0.2350). It indicates the amount of the preference for housing (leisure) that is not explained by individual observables is minor relative to what the covariates explain.

15 Predicted Outcomes

In order to test how well the model performs, I compare three types of predicted, aggregate moments, evaluated at the estimated values of the parameters, $\hat{\theta}$, with their real-world counterparts from the data. Each of the moments corresponds to one of the modeled outcomes: commute time, commute mode, or housing choice. I begin with commute time, as it is the simplest outcome to calculate and is an input into other predicted outcomes. Next, I present moments relating to commute mode choice and finally housing choice.

To clarify the notation used in this section, recall that the observed housing choice of family i is h, the observed commuting method of the head of family i is k, and the head of family i commutes for a duration of t_{ihk}. I define $\hat{h}$ as a possible home (from the affordable choice set of homes) family i could live in, $\hat{k}$ as a potential method (again, from the feasible choice set) the individual i could use to get to work, and $\hat{t}_{i\hat{h}\hat{k}}$ as the predicted time it would take her to get to work from the given home by the given method.

15.1 Commute Time

To measure how well the model predicts commute times, I compare predicted commute times to the distribution of actual commute times from the data. Recall from Section B.1.1 that $t_{ihk} = \exp\left(K_{ihk}\gamma_k^K + e_{ihk}\right)$, so the predicted commute time is $\hat{t}_{i\hat{h}\hat{k}} = \int \exp\left(K_{i\hat{h}\hat{k}}\hat{\gamma}_{\hat{k}}^K + e_{i\hat{h}\hat{k}}\right) dF\left(e_{i\hat{h}\hat{k}}\right)$. The commute characteristic parameters, γ^K, are estimated in the P_i^t equation which uses the observed commute time as the response variable. The γ_k^K are identified from the covariation of the commute characteristics (K_{ihk}) and that response variable, but they also appear elsewhere in the model. They are used in the P_i^{HK} equation to proxy for unobserved commute times, so accurate estimates of commute times are an important input into the residential and commute mode choice parts of the model. I am in the process of calculating these measures and will present and analyze them as part of the published version of this research project.

15.2 Commute Mode Choice

The probability an individual chooses a given $\left(\hat{h},\hat{k}\right)$ pair is the predicted analog to Equation 7, the joint probability of observing a family living in house h and commuting by method k. I define this predicted probability as $\hat{P}^{HK}_{i\hat{h}\hat{k}} = \Pr(\hat{h},\hat{k} \mid \hat{\theta},\hat{t}_{i\hat{h}\hat{k}})$.[54] Summing this probability over homes for a given $\hat{k}$ gives the predicted probability that the commuter from family i commutes by method $\hat{k}$: $\hat{P}^{K}_{i\hat{k}} = \Pr(\hat{k} \mid \hat{\theta},\hat{t}_{i\hat{h}\hat{k}}) = \sum_{\hat{h}=1}^{N^H} \hat{P}^{HK}_{i\hat{h}\hat{k}}$. I take the mean of these probabilities over individuals by commute mode to calculate the average predicted probability of commuting by the given method. Formally, I calculate the average predicted probabilities as $\bar{P}^{K}_{\hat{k}} = \frac{1}{N^I}\sum_{i=1}^{N^I} \hat{P}^{K}_{i\hat{k}}\,\forall\,\hat{k}$. I am in the process of calculating these probabilities and will present and analyze them as part of the published version of this research project.

15.3 Housing Choice

As analogous aggregate measures to the ones detailed for commute mode choices (that aggregate $\hat{P}^{HK}_{i\hat{h}\hat{k}}$ to show the average probability each individual lives in each home) are not meaningful because each home can house only one family, so I aggregate in a different manner. I am interested in how policies that affect commutes influence the distribution of housing locations, so I use geographic information on housing and job locations to calculate the aggregate probability that individuals live within a given distance range from their work locations. I index these ranges with l and define the bounds of these ranges as $\left(\overleftarrow{d_l}, \overrightarrow{d_l}\right)$. Recall from Section 4.3 that d_{ihk} is the distance between house h and job j that individual i travels by commuting method k. I define an indicator function that is equal to 1 if $d_{\hat{h}j}$ falls within range l: $1\left(\overleftarrow{d_l} < d_{\hat{h}j} < \overrightarrow{d_l}\right)$. Finally, I can write the individual probability of interest as $\hat{P}^{H}_{il} = \Pr(\overleftarrow{d_l} < d_{\hat{h}j} < \overrightarrow{d_l} \mid \hat{\theta},\hat{t}_{i\hat{h}\hat{k}}) = \sum_{\hat{h}=1}^{N^H}\sum_{\hat{k}=1}^{N^K} \hat{P}^{HK}_{i\hat{h}\hat{k}} 1\left(\overleftarrow{d_l} < d_{\hat{h}j} < \overrightarrow{d_l}\right)\,\forall\,l$, and it's aggregate analog as $\bar{P}^{H}_{l} = \frac{1}{N^I}\sum_{i=1}^{N^I} \hat{P}^{H}_{i\hat{d}}\,\forall\,l$. I am in the process of calculating these probabilities, as well as developing other measures of interest, for the published version of this research project.

16 Specification Tests

To assess the accuracy of the model, I conduct a several specification tests. In this section, I explain how I conduct a chi-square goodness-of-fit test to asses how well the model performs. I am also working to develop additional specification tests including Wald tests of whether relevant subsets of the parameters are jointly equal to zero and Lagrange Multiplier tests to confirm that the model is properly specified.

16.1 Chi-Square Goodness-of-Fit Test

I use chi-square goodness-of-fit tests to determine how well the model reflects the data. I perform two tests that relate to the previously outlined outcome probabilities. First, I test the null hypothesis that the observed and predicted proportion individuals commuting by each mode are identical.

[54] Note that this notation means that I use the predicted value of $\hat{t}_{ihk}$, not the observed commute time, in calculating the predicted probability of observing (h,k). For all other $\left(\hat{h},\hat{k}\right)$, this is the only option.

Formally, this test is

$$\begin{aligned} H_0 : \bar{P}_k^K N^I &= \bar{P}_{\hat{k}}^K N^I \\ H_A : \bar{P}_k^K N^I &\neq \bar{P}_{\hat{k}}^K N^I \, for\, each\, k. \end{aligned}$$

The χ^2 statistic for each commuting method is

$$\frac{\left[N^I \left(\bar{P}_k^K N^I - \bar{P}_{\hat{k}}^K N^I\right)\right]^2}{\bar{P}_{\hat{k}}^K N^I},$$

which has a χ^2 distribution with $N^K - 1$ degrees of freedom.

I also perform a similar test on the observed and predicted average probability of living within a given distance range from work. The test and χ^2 statistic are defined similarly to the previous case as

$$\begin{aligned} H_0 : \bar{P}_l^H \left(d_{hj}\right) N^I &= \bar{P}_l^H \left(d_{\hat{h}j}\right) N^I \\ H_A : \bar{P}_l^H \left(d_{hj}\right) N^I &\neq \bar{P}_l^H \left(d_{\hat{h}j}\right) N^I \, for\, each\, l, \end{aligned}$$

and

$$\frac{\left[N^I \left(\bar{P}_l^H \left(d_{hj}\right) - \bar{P}_l^H \left(d_{\hat{h}j}\right)\right)\right]^2}{\bar{P}_l^H \left(d_{\hat{h}j}\right) N^I}.$$

Letting N^d denote the number of distance ranges indexed by l, the χ^2 statistic follows a distribution with $N^d - 1$ degrees of freedom. As I have not yet calculated the predicted probabilities, I am unable to present the results of these tests, however they will be completed as part of the published version of the research project.

17 Policy Simulations

Congestion is the result of the nature of impure public goods (roads are non-excludable, but are rival) that causes them to be provided by the government at zero marginal cost. The rivalry leads to external costs because each additional driver on the road imposes costs on her fellow commuters that she does not fully bear. Direct quotas and Pigouvian taxes on vehicle miles traveled during congested times of the day are politically infeasible first best solutions (Parry et al. (2007)). Congestion pricing has been gaining traction as a more feasible alternative means of reducing congestion, and Shoup (1997) advocates applying congestion pricing principles to public street parking to reduce congestion, amongst other benefits.[55] All of these policies have the potential to influence both the monetary and time costs of commuting. I seek to better inform the discussion of

[55] I explain the details of several congestion pricing policies in Section 3.

ways to reduce congestion by performing simulations that illuminate the response to shifts in costs caused by a given policy. My model allows me to account for the response to policy shifts both in terms of the distribution of commuting method and residential location decisions.

I am in the process of using my model estimates to performing these comparative statics. I plan to conduct policy experiments based on numerous policies. Based on the parameters of a given policy, I first alter the pecuniary and time costs of the commutes faced by individuals in my model. Then I allow for three types of responses. First, I allow for a short-term response in terms of mode choice only. Second, I allow individuals to switch to a different commuting mode and/or move to a new residence in the medium term. Finally, in the long term, I also allow the housing stock to respond. The first two responses require only that I include a measure of congestion that feeds-back individual responses into the model.[56] The later requires estimating an additional housing stock equation. Although I do not estimate an equilibrium model, I can perform equilibrium comparative statics using the following algorithm:

1. Change the cost inputs based on the parameters of the given policy,

2. Calculate the distribution of k that the model predicts with the new costs (do the same for h in the later two scenarios) ,

3. Recalculate t_{ihk} based on the new distribution of k (and h, where applicable),

4. Recalculate p^H based on the new distributions of k and h (in the later two scenarios), then

5. Repeating the previous steps until the process converges to a state where individuals no longer change their commuting mode or housing location.

The results of these simulations will allow me to determine not only the effects of a given policy on congestion, but also how much of that effect is due to mode switching and how much is due to individuals moving to new residential locations. I can also analyze which individuals are affected by the given policy to determine whether the policy is regressive in nature. This is an important consideration that is often cited by opponents of congestion pricing policies (Parry et al. 2007, Lewis 2008).

Part VII

Conclusions

My research develops a structural model of residential choice and commuting that makes contributions to both the transportation and residential choice literatures. I do so by addressing the en-

[56] I have experimented with aggregate measures of congestion to capture this effect, but they are not identified by time variation alone with only one market included in my estimation. Instead, I plan to develop a measure of congestion based on the number of individuals commuting from an area around a given home to an area around a given individual's job location.

dogeneity of residential choice in analysis of commuting behavior with an individual-level model that has a rich unobserved heterogeneity structure. The restricted-access data used for estimation contains geographic precision that allows me to use GIS network analysis to painstakingly model the optimal commute between each pairwise combination of home and job locations by each commuting method observed in the data. The model works well in predicting behavior, as evidenced by the reasonable, preliminary estimates presented. Finally, I outline policy simulations that are directly relevant to an emerging policy, the effects of which we do not yet fully understand, that has the potential to drastically reshape the urban environment in this country. Future work will focus on improving the fit of my model and conducting the policy simulations I outline. A follow up project that is currently in the work-in-progress stage will add cohabiting couples to my estimation routine.

References

Alonso, William (1964) *Location and Land Use: Toward a General Theory of Land Rent*: Cambridge, Mass.: Harvard Univ. Pr. 5

Anas, Alex and Robin Lindsey (2011) "Reducing Urban Road Transportation Externalities: Road Pricing in Theory and in Practice," *Review of Environmental Economics and Policy*, Vol. 5, pp. 66–88, URL: `http://reep.oxfordjournals.org/content/5/1/66.abstract`. 7

Arévalo, Raquel and Javier Ruiz-Castillo (2006) "On the Imputation of Rental Prices to Owner-Occupied Housing," *Journal of the European Economic Association*, Vol. 4, pp. 830–861, URL: `http://dx.doi.org/10.1162/JEEA.2006.4.4.830`, DOI: `http://dx.doi.org/10.1162/JEEA.2006.4.4.830`. 46

Bajari, Patrick and Matthew E Kahn (2005) "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities," *Journal of Business & Economic Statistics*, Vol. 23, pp. 20–33, URL: `http://amstat.tandfonline.com/doi/abs/10.1198/073500104000000334`, DOI: `http://dx.doi.org/10.1198/073500104000000334`. 3, 5

Bajari, Patrick and Matthew E. Kahn (2008) "Estimating Hedonic Models of Consumer Demand with an Application to Urban Sprawl," in Andrea Baranzini, José Ramirez, Caroline Schaerer, and Philippe Thalmann eds. *Hedonic Methods in Housing Markets*: Springer New York, pp. 129–155, URL: `http://dx.doi.org/10.1007/978-0-387-76815-1_7`. 3, 5

Banzhaf, H. Spencer and Omar Farooque (2012) "Interjurisdictional Housing Prices and Spatial Amenities: Which Measures of Housing Prices Reflect Local Public Goods?" Working Paper 17809, National Bureau of Economic Research. 45

Baum-Snow, N. and M.E. Kahn (2000) "The Effects of New Public Projects to Expand Urban Rail Transit," *Journal of Public Economics*, Vol. 77, pp. 241–263. 2, 4, 28

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007) "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, Vol. 115, pp. 588–638, URL: `http://ideas.repec.org/a/ucp/jpolec/v115y2007p588-638.html`. 9, 23, 25, 44, 45, 48

Bayer, Patrick and Robert McMillan (2012) "Tiebout sorting and neighborhood stratification," *Journal of Public Economics*, Vol. 96, pp. 1129–1143, URL: `https://ideas.repec.org/a/eee/pubeco/v96y2012i11p1129-1143.html`. 3, 5

Bayer, Patrick, Robert McMillan, and Kim Rueben (2005) "An Equilibrium Model of Sorting in an Urban Housing Market," Working Paper 10865, National Bureau of Economic Research. 8, 23, 44, 48

Bayer, Patrick, Robert McMillan, and Kim S. Rueben (2004) "What drives racial segregation? New evidence using Census microdata," *Journal of Urban Economics*, Vol. 56, pp. 514–535, URL: `http://ideas.repec.org/a/eee/juecon/v56y2004i3p514-535.html`. 5, 25

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa (2008) "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, Vol. 116, pp. 1150–1196, URL: `http://ideas.repec.org/a/ucp/jpolec/v116y2008i6p1150-1196.html`. 5, 25

Bento, Antonio M., Maureen L. Cropper, Ahmed Mushfiq Mobarak, and Katja Vinha (2005) "The Effects of Urban Spatial Structure on Travel Demand in the United States," *Review of Economics and Statistics*, Vol. 87, pp. 466–478, URL: `http://www.mitpressjournals.org/doi/abs/10.1162/0034653054638292`, DOI: `http://dx.doi.org/10.1162/0034653054638292`. 2, 4

Berndt, E.K., B.H. Hall, and R.E. Hall (1974) "Estimation and Inference in Nonlinear Structural Models," in *Annals of Economic and Social Measurement, Volume 3, number 4*: NBER Books, pp. 103–116. 27

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, Vol. 63, pp. pp. 841–890, URL: `http://www.jstor.org/stable/2171802`. 5

Bunch, David S. (1991) "Estimability in the multinomial probit model," *Transportation Research Part B: Methodological*, Vol. 25, pp. 1 – 12, URL: `http://www.sciencedirect.com/science/article/pii/0191261591900098`, DOI: `http://dx.doi.org/10.1016/0191-2615(91)90009-8`. 54

Dettling, Lisa J. and Melissa Schettini Kearney (2011) "House Prices and Birth Rates: The Impact of the Real Estate Market on the Decision to Have a Baby," Working Paper 17485, National Bureau of Economic Research. 56

Duranton, Gilles and Matthew A. Turner (2009) "The Fundamental Law of Road Congestion: Evidence from US cities," NBER Working Papers 15376, National Bureau of Economic Research, Inc. 4

——— (2011) "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, Vol. 101, pp. 2616–52, URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2616, DOI: http://dx.doi.org/10.1257/aer.101.6.2616. 2

Eliasson, Jonas (2014) "The Stockholm congestion pricing syndrome: how congestion charges went from unthinkable to uncontroversial," Working papers in Transport Economics 2014:1, CTS - Centre for Transport Studies Stockholm (KTH and VTI). 2

Geweke, John (1989) "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, Vol. 57, pp. pp. 1317–1339, URL: http://www.jstor.org/stable/1913710. 23

Goodman, Allen C. and Masahiro Kawai (1985) "Length-of-Residence Discounts and Rental Housing Demand: Theory and Evidence," *Land Economics*, Vol. 61, pp. pp. 93–105, URL: http://www.jstor.org/stable/3145802. 46

Goodman Jr., John L. and John B. Ittner (1992) "The accuracy of home owners' estimates of house value," *Journal of Housing Economics*, Vol. 2, pp. 339 – 357, URL: http://www.sciencedirect.com/science/article/pii/105113779290008E, DOI: http://dx.doi.org/10.1016/1051-1377(92)90008-E. 45

Hajivassiliou, Vassilis A. (1990) "Smooth Simulation Estimation of Panel Data LDV Models," Unpub. msc. Yale U., 1990. 23

Houde, Jean-François (2012) "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, Vol. 102, pp. 2147–82, URL: http://www.aeaweb.org/articles_detail.php?doi=10.1257/aer.102.5.2147. 20

Ioannides, Yannis M. (2012) *From Neighborhoods to Nations*: Princeton University Press, pp.552, URL: http://www.jstor.org/stable/j.cttq95gx. 5

Keane, Michael P. (1994) "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, Vol. 62, pp. pp. 95–116, URL: http://www.jstor.org/stable/2951477. 23

Kiel, Katherine A. and Jeffrey E. Zabel (1999) "The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey," *SSRN eLibrary*, Vol. 27. 45

——— (2008) "Location, location, location: The 3L Approach to house price determination," *Journal of Housing Economics*, Vol. 17, pp. 175 – 190, URL: http://www.sciencedirect.com/science/article/pii/S105113770800003X, DOI: http://dx.doi.org/10.1016/j.jhe.2007.12.002. 25

Krueger, Alan B., Daniel Kahneman, David Schkade, Norbert Schwarz, and Arthur A. Stone (2008) "National Time Accounting: The Currency of Life," Working Papers 523, Princeton University, Department of Economics, Center for Economic Policy Studies. 9

Langer, Ashley and Clifford Winston (2008) *Brookings-Wharton Papers on Urban Affairs 2008*, Chap. Toward a Comprehensive Assessment of Road Pricing Accounting for Land Use, pp. 127–175: The Brookings Institution Press, URL: http://www.brookings.edu/research/papers/2008/12/road-pricing-winston. 3, 5, 57

Leape, Jonathan (2006) "The London Congestion Charge," *The Journal of Economic Perspectives*, Vol. 20, pp. pp. 157–176, URL: http://www.jstor.org/stable/30033688. 2, 6

Lerman, S.R. (1976) "Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model," *Transportation Research Record*, Vol. 610, pp. 6–11. 5

Lewis, David (2008) "America's Traffic Congestion Problem: Toward a Framework for Nationwide Reform," *Brookings Institute Discussion Paper*, Vol. 2008-06. 2, 6, 37

Lindsey, R. (2006) "Do Economists Reach a Conclusion?" *Econ Journal Watch*, Vol. 3, pp. 292–379. 6

Malpezzi, Stephen (2008) *Hedonic Pricing Models: A Selective and Applied Review*, Chap. 5, pp. 67–89: Blackwell Science Ltd, URL: http://dx.doi.org/10.1002/9780470690680.ch5, DOI: http://dx.doi.org/10.1002/9780470690680.ch5. 49

Marshall, Robert C. and J. Luis Guasch (1983) "OCCUPANCY DISCOUNTS IN THE U.S. RENTAL HOUSING MARKET*," *Oxford Bulletin of Economics and Statistics*, Vol. 45, pp. 357–378, URL: http://dx.doi.org/10.1111/j.1468-0084.1983.mp45004003.x, DOI: http://dx.doi.org/10.1111/j.1468-0084.1983.mp45004003.x. 46

McFadden, D. (1974) *Conditional Logit Analysis of Qualitative Choice Behavior*, pp. 105–142: Academic Press: New York. 4, 23

McFadden, Daniel (1978) "Modelling the Choice of Residential Location," in A Karlqvist, L Lundqvist, F Snickars, and J Weibull eds. *Spatial Interaction Theory and Planning Models*, Vol. 3, Amsterdam: North Holland, pp. 75–96. 5, 16

——— (2001) “Disaggregate Behavioral Travel Demand’s RUM Side: A 30-Year Retrospective,” in David Heshner ed. *The Leading Edge of Travel Behavior Research*, Oxford: Pergamon Press. 4, 8

Mills, Edwin S. (1967) “An Aggregative Model of Resource Allocation in a Metropolitan Area,” *The American Economic Review*, Vol. 57, pp. pp. 197–210, URL: `http://www.jstor.org/stable/1821621`. 5

Minnesota Population Center (2011) “National Historical Geographic Information System: Version 2.0,” Minneapolis, MN: University of Minnesota, URL: `http://www.nhgis.org`.

Muth, Richard (1969) *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*: Chicago: The University of Chicago Press. 5

Olsen, Edgar O., Dirk W. Early, and Paul E. Carrillo (2012) “A Panel of Price Indices for Housing Services, Other Goods, and All Goods for All Areas in the United States 1982-2010,” Virginia Economics Online Papers 402, University of Virginia, Department of Economics. 23

Parry, Ian W. H., Margaret Walls, and Winston Harrington (2007) “Automobile Externalities and Policies,” *Journal of Economic Literature*, Vol. 45, pp. 373–399, URL: `http://ideas.repec.org/a/aea/jeclit/v45y2007i2p373-399.html`. 2, 6, 36, 37

Pisarski, Alan (2006) “Commuting in America III: The Third National Report on Commuting Patterns and Trends,”Technical report, Transportation Research Board. 43

Polzin, Steven and Xuehao Chu (2005) “Public Transit in America: Results from the 2001 National Household Travel Survey,”Technical report, National Center for Transit Research at the Center for Urban Transportation Research at the University of South Florida. 43

Reschovsky, Clara (2004) “Journey to Work: 2000,”Technical Report C2KBR-33, United States Census Bureau. 43

Schrank, David, Bill Eisele, Tim Lomax, and Jim Bak (2015) “The 2015 Urban Mobility Scorecard,”Technical report, Texas A&M Transportation Institute and INRIX, Inc. Texas Transportation Institute, Texas A&M University System and INRIX, Inc. 2, 13

Shoup, Donald C. (1997) “The High Cost of Free Parking,” *Journal of Planning Education and Research*, Vol. 17, pp. 3–20, URL: `http://jpe.sagepub.com/content/17/1/3.abstract`. 36

Small, Kenneth A. and Erik T. Verhoef (2007) *The Economics of Urban Transportation*: Routledge, 2nd edition. 4

Small, Kenneth A., Clifford Winston, and Jia Yan (2005) “Uncovering the Distribution of Motorists’ Preferences for Travel Time and Reliability,” *Econometrica*, Vol. 73, pp. 1367–1382, URL: `http://dx.doi.org/10.1111/j.1468-0262.2005.00619.x`. 7

——— (2006) "Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design," *Brookings-Wharton Papers on Urban Affairs*, pp. pp. 53–96, URL: `http://www.jstor.org/stable/25067428`. 7

Stern, Steven (1997) "Simulation-Based Estimation," *Journal of Economic Literature*, Vol. 35, pp. 2006–2039, URL: `http://www.jstor.org/stable/2729885`. 23, 27

Tiebout, Charles M. (1956) "A Pure Theory of Local Expenditures," *Journal of Political Economy*, Vol. 64, pp. 416–424, URL: `http://www.jstor.org/stable/1826343`. 5

Train, K.E. (2009) *Discrete Choice Methods with Simulation*: Cambridge University Press, 2nd edition. 4, 23, 55

Vega, Amaya and Aisling Reynolds-Feighan (2009) "A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns," *Transportation Research Part A: Policy and Practice*, Vol. 43, pp. 401 – 419, URL: `http://www.sciencedirect.com/science/article/pii/S0965856408002103`. 6, 16

Part VIII
Appendix

A Data

A.1 Census Commuting Questions

Pisarski (2006) provides a list of shortcomings in the Census journey-to-work data that begins with the fact that the data contains no information "about aspects of trips using more than one mode of travel to get to work." According to the Census 2000 Brief "Journey to Work: 2000," Census data report the "usual means of transportation to work." When a person usually commutes via multiple transportation methods, only the method that covers the greatest distance is recorded (Reschovsky 2004). Given the prevalence of park-and-rides, transfer passes, and bike racks on buses, it is clear that multimodal travel is a reality in modern commuting, but there is little data available on this type of behavior. The National Household Travel Survey (NHTS) is the only national survey that measures mode transfers. Polzin and Chu (2005) calculate that 20 percent of all daily travel trips on transit are multimodal based on NHTS data, but the authors cannot reconcile this estimate of the prevalence of transfers with aggregate counts of the number of individuals who board transit vehicles reported by the NTD and the American Public Transportation Association (APTA). This discrepancy suggests that the NHTS figure may be an underestimate.[57]

While my research does not address this concern directly as I am not able to identify the exact secondary means of transportation to work, I can mitigate the effects of unobserved commuting methods. I do so in two ways. First, I restrict the sample to by dropping individuals who are obvious multimodal commuters. For instance, an individual who reports walking as her primary means of commuting, but lives many miles from her job. Second, I calculate the as-the-crow-flies distance from the fixed locations where individuals can enter and exit transit systems to their home and office locations. I assume that all commuters travel to the closest transit station to their home and exit at the closest transit station to their work location when traveling to work. To account for the fact that, for instance, those who live close to a station most likely walk or bike there, and beyond some threshold distance, individuals likely take the bus or drive to the station, I allow for differential effects by distance. Unfortunately, this control only works for multimodal commuters who report a form of rail as their primary means of commuting. I cannot apply a similar control to those who report commuting by road as their primary means of commuting.

A.2 Opportunity Cost of Housing

The ACS includes two types of housing costs depending on the tenure type of the family being surveyed (ie, home owner or renter). Home-owners self-report a measure of total property value and renters report their monthly rent. Bayer et al. (2005) and Bayer et al. (2007) explain that there are three concerns with interpreting this data as a continuous measure of the opportunity cost of living in the given home. First, both the property value and rent variables may not reflect the true market value of the home. For home owners, the likely culprits are misreporting and overestimating the value of one's home. The real estate market is fluid and keeping up with it is costly, so home owners may not be savvy to the current market value of their home if they did not purchase it in the recent past (or if they have no intention of selling it in the near future). They may also have a more optimistic outlook on the value of their home than is warranted. While renters are much more likely to know their monthly rent, that rent may reflect a tenure discount if they have lived in the home for an extended period of time. A second issue is that, property values are reported in intervals in surveys prior to 2008 and are top-coded in all years. Finally, one must also adjust the owner and renter home value measures to be compatible across tenure types, as home values reflect the present discounted value of the flow of value from the home and rents reflect the stock value of the home.

This section details how I account for these issues and construct a consistent measure of the opportunity cost of living in each home from the available data. I primarily build on the data cleaning procedures of the Bayer et al. papers, however, since both studies use data from the 1990 Decennial Census for the San Francisco Bay Area, I modify their methodology to more appropriately fit my model and data. While the questions asked in recent Decennial Censuses and the ACS are remarkably similar, my data differs in three key ways. First, my data is a repeated cross

[57] Note that the NHTS figure reports the number of transfers on daily travel trips of all types, not just trips made commuting to and from work. It is not clear whether controlling for this distinction biases the estimate up or down.

section that spans multiple years as opposed to just one. Second, my data pertains to a different metropolitan area. This is noteworthy because of institutional differences in the way property taxes are assessed. Finally, property values are reported categorically in Census products prior to 2008 and continuously thereafter.

I proceed by detailing adjustments made to property values, rent values, and tenure type.

A.2.1 Property Value

Home-owners are asked to self-report the value of their home and property and this data is reported as a categorical, top-coded variable. Bayer et al. (2007) find that owners frequently report their home's purchase price, not its current value. There is evidence that this effect is present in my data as well. Homes sold within the previous year have, on average, reported values that are 10% higher than observationally equivalent homes purchased between 20 and 30 years earlier, all else equal.[58] In addition to misreporting, it has been shown in the literature that home owners frequently overestimate the value of their homes using comparisons of self-reported and housing transactions data (see, for instance, Goodman Jr. and Ittner (1992), Kiel and Zabel (1999)). While I cannot determine the prevalence of overestimation of home prices due to the lack of transactions data, Banzhaf and Farooque (2012) find that price indices based on self-reported home values are highly correlated with those based on transactions data and are a practical alternative to more accurate, but less available, transactions data. To correct for the differential effects of misreporting across different categories of the family's tenure in the home and account for the overestimation of home values in my self-reported data, I estimate a house value hedonic at the community level using interval regression and use this regression to predict a continuous variable from the categorical, top-coded data. Doing so at the community level is equivalent to computing a price index (see Banzhaf and Farooque (2012), footnote 10), so this measure should perform as well as one based on unavailable, but more precise transactions data.

Formally, I interval regress log home value on tenure categories, annual property taxes paid, their interactions, housing characteristics, and year indicators. I do so separately by Public Use Microdata Areas (PUMAs).[59] I estimate the following equation

$$\ln(V_h) \quad = \quad \alpha_1 tenure_h + \alpha_2 \ln(tax_h) + \alpha_3 (tenure_h \times \ln(tax_h)) + \alpha_4 H_h + \alpha_5 year_h + \omega_h^V,$$

where V_h is the self-reported house value, $tenure_h$ is a categorical measure of the length of time the family has resided in home h, tax_h is the self reported property taxes paid, H_h is the set of housing characteristics, $year_h$ is an indicator for the year the data was collected, and ω_h^V is an error. Bayer et al. (2007) are able to use the rules associated with Proposition 13 to transform property taxes paid into an estimate of the home's current value. I depart from their framework by including property taxes paid instead of this estimate, which I cannot easily calculate because property tax

[58] All calculations reported in this appendix are based on 2005-2008 ACS PUMS data.

[59] Estimates based on regressions at the PUMA and year level did not substantially improve results.

laws vary over time and with geography in my sample.[60] However, since I am running separate regressions at the PUMA level, tax laws should be close to consistent by regression, although rates will undoubtedly vary over time. If this is the case, α_2 will have predictive power so long as homes that have higher property taxes have higher values and it will return a linear approximation to the property tax cost in the PUMA. To the extent that multiple jurisdictions may exist in a given PUMA, α_2 will return a weighted average of these costs. To reduce the influence that misreporting associated with longer tenured homes has on the fitted values of the hedonic, I interact the tenure and property tax rates. Finally, I replace V_h with $\hat{V}_h$ in subsequent steps to correct for misreporting.

A.2.2 Rental Value

The existence of substantial tenure discounts in the rents of residents based on their length-of-residence in a given home is a well known phenomenon in the literature. For example, Marshall and Guasch (1983) are unable to reject the existence of such discounts. Goodman and Kawai (1985) find that the rent of recent movers is between 4% and 11% greater than that of all renters, depending on specification. More recently, Arévalo and Ruiz-Castillo (2006) report discounts in Spanish housing markets ranging from 3.2% to 83.5%, depending on the length-of-residence (up to 25 years). Discounts in line with these estimates exist in my data: renters who are in the second year of their lease receive a 4% discount relative to renters in the first year of their lease, all else equal. This discount increases to 50% for individuals who have lived in their residence for between 20 and 30 years.

Tenure discounts are believed to be the result of unobserved heterogeneity due to depreciation and/or state dependence due to match quality between the landlord and the tenant. The first explanation posits that if landlords postpone performing maintenance or reconditioning a home until tenant turnover, homes with longstanding tenants will be of lower quality than those available in the market. To the extent that the available information in the data does not accurately measure the quality of a unit (for instance, the data provides the number of bedrooms, but not how recently the carpet in those bedrooms was replaced), this depreciation will be unobserved and explains the existence of a tenure discount as a means of accounting for quality differences. An alternative explanation is due to state dependence. Arévalo and Ruiz-Castillo (2006) explain that turnover is costly not only for the tenant, but also the landlord (the costs of filling a vacancy include advertising costs, forgone rent, etc.). In addition, landlords may want to retain "good" tenants who treat the unit well and coexist with their neighbors. Landlords may do so by offering a discount to tenants who reveal themselves to be of high quality (see Goodman and Kawai (1985) for a theoretical model). Note that it is not possible to determine which phenomena leads to tenure discounts and both are likely to play a role in their existence.[61]

[60] Washington, DC assess property taxes at the district level. Virginia assess property taxes at the county, city, or town level, and Maryland assess property taxes at the county or city level.

[61] Ideally, I would be able to model home choice as a dynamic programing problem where individuals choose their optimal home in each period given their expectations about future utility flows from the home (net of ownership or rental costs). With multiple observations on renters and homes, I would be able to separately identify the cause of the tenure discount and adjust the rent each family would face at each home accordingly. Unfortunately, the data preclude

These discounts are of consequence when I construct a measure of market rent for every home in my sample. Doing so requires capturing the unmodeled dynamics that generate tenure discounts. The salient question is: what rent would a family in the model pay in each home *other* than the one the family lives in? There are four ways to construct a measure of unobserved rents, either as

1. The reported rent in the given home, which includes any tenure discounts that the current tenant has accrued or

2. An estimate of the rent in the given home that excludes the current tenant's tenure discount and instead includes an estimated tenure discount based on how long the family has lived in its observed home or

3. An estimate of the rent in the given home that excludes all tenure discounts or

4. An estimate of the rent in the given home that includes an estimated tenure discount based on what the current tenant has accrued.

How one interprets the cause of tenure discounts can help to guide the decision of which method is best, but they all have their drawbacks.

The first method of constructing a measure of unobserved rents is to naively assume that the observed rent in the given home is the market rent. This is not sensible because it implies strong assumptions about the nature of both sources of tenure discounts. If the discount is entirely due to unobserved depreciation, using the observed rent without adjusting for duration of tenure implicitly assumes that landlords do not perform maintenance on the apartment before new tenants move in. However, this assumption negates the explanation for why unobserved depreciation results in a tenure discount that is not captured by controlling for the age of the housing structure. If, on the other hand, the discounts are entirely due to state dependence, this method of construction assumes that the discount is solely the result of the characteristics of the landlord (not the quality of the match between the tenant and landlord). Again, this negates the explanation for why state dependence results in a tenure discount: the landlord would just offer a low rent to new tenants to quickly fill his apartment if he was unconcerned with match quality.

The second means of constructing the rent measure equalizes the family's discount across all homes. The method is paramount to assuming that the family, however many years ago it was searching for its current home, faced the options that currently exist in the data and made a decision about where to live. The benefit of "turning back the clock" in this way is that doing so not imply any assumptions about the nature of the unobserved depreciation or state dependence that generates the discounts. However, this method is difficult to implement in practice because it would require adjusting the time dependent observable characteristics of the family members (such as age and marital status) back to what they were when the family last moved. Additionally, it is not an accurate representation of the decision a family thinking about moving in the given period faces, as it assumes that the family has perfect foresight and decides where to move once and stays there.

a dynamic model, as they are cross-sectional in nature.

The bias that using this method would introduce into my model depends on the reason tenure discounts exist. To the extent that tenure discounts are due to unobserved depreciation (and the discount prices this depreciation appropriately), the family would be indifferent between moving to a higher cost, recently maintained home and staying in their depreciated home with a discount, so no bias would be introduced. To the extent that tenure discounts are due to state dependence, families in my model would be more apt to move because they would not forfeit their accrued tenure discount by moving. If this is the case, this method would overstate the response to a policy that shifts the distribution of housing.

The third method of constructing the rent measure removes tenure discounts from all homes. It is equivalent to assuming that the family is living in the current home and considering moving to the other homes in the choice set. Whether consciously or not, this is a choice that families make each period. If the cause of tenure discounts is entirely due to unobserved depreciation, this means of constructing the rent measure assumes that landlords perform maintenance on homes before new tenants move in and adjust rents accordingly, so it is consistent with the unobserved depreciation theory. If discounts exist because of state dependence, this method assumes that they are the result of a good match between the both the tenant and landlord, which cannot be known to either the landlord or the tenant when they first sign a lease. This method is also consistent with the state dependence theory. While all of these implications are reasonable, the drawback of this method is that the family's housing history is endogenous because the discount is not removed from the home the family is currently living in.[62] This means that my model would understate a family's willingness to move in response to a policy shift because doing so would mean forfeiting an accrued tenure discount. Again, to the extent that tenure discounts are due to appropriately priced depreciation, this concern would be mitigated because the family would be indifferent between moving to a higher cost, recently maintained home and staying in their depreciated home with a discount. However, if the discount is caused by state dependence, this issue would be of greater concern.

The fourth method is similar to the first, but smooths the tenure discount by using the aggregate market discount instead of the individual home/landlord/renter discount. As all four methods of construction have drawbacks, I proceed by following the fourth method because it mitigates the extreme assumptions of other methods, it can be implemented without requiring ad-hoc adjustments to the time dependent observable characteristics of the family members, and it most closely reflects the standard in the literature set forth by Bayer et al. (2005) and Bayer et al. (2007). To model the market rents as the rent associated with the home reported in the data after smoothing with an aggregate measure of the tenure discount, I regress

$$\ln(R_h) = \beta_1 tenure_h + \beta_2 H_h + \beta_3 year_h + \omega_h^R,$$

where R_h is the self reported gross rent, $tenure_h$, $year_h$, and H_h are as described in the home value

[62]This problem could be ameliorated by adjusting rents for tenants in their observed homes to reflect a market rent that excludes tenure discounts, but doing so would mean that some individuals would not be able to afford the home they live in.

equation, and ω_h^R is an error.[63] Again, I run separate regressions by PUMA. I use $\hat{R}_h$ as a measure of the market rent in the given home.

A.2.3 Adjustment for Tenure Type

Data on owner and renter home costs in the data are not compatible because home values represent the present discounted value of a flow of services from the home and rents are the stock value of those services. In order to calculate the opportunity cost of living in an owner-occupied home, I regress

$$\ln(\pi_h) = \gamma_1 o_h + \gamma_2 H_h + \gamma_3 year_h + \omega_h^p,$$

by PUMA, where

$$\pi_h = \begin{cases} \hat{V}_h & if\, o_h = 1 \\ \hat{R}_h & else, \end{cases}$$

$$o_h = \begin{cases} 1 & if\, h\, is\, owner-occupied \\ 0 & else, \end{cases}$$

H_h and $year_h$ are as described in the previous equations, and ω_h^p is an error. I then use the estimate of γ_1 to convert home values to a measure of the rent the family would pay if it were renting the home. After making these adjustments to the data, the price of housing as defined in my model, is

$$p_h^H = \exp(\hat{\pi}_h(o_h = 0)).$$

B Estimation

B.1 Joint Probability of Observing t_{ihk}, h, and k

B.1.1 Probability of Observing t_{ihk}

Recall from Equation 1 that the observed commute time is a function of both a linear index of the characteristics of individual i's commute and an error term. The probability that the individual's observed commuting time is equal to the commuting time the model predicts is the probability that this equality holds for the observed home and commuting method:

$$P_i^t = \Pr\left(t_{ihk} = \exp\left(K_{ihk}\gamma_k^K + e_{ihk}\right)\, for\, h\, and\, k \mid \theta\right).$$

Explicitly, this is

[63] To be consistent with home values, I use contract rents instead of gross rents (gross includes applicable home utility costs such as heat, electricity, etc.). In order to account for the effects of the inclusion of utilities in rental costs in some homes, but not in others, I follow Malpezzi (2008) and include indicators for the inclusion of a given utility in the rent in H_h.

$$P_i^t = \frac{1}{\sigma_e t_{ihk}} \phi \left(\frac{\ln(t_{ihk}) - K_{ihk}\gamma_k^K}{\sigma_e} \right),$$

where $\phi(\cdot)$ is the standard normal probability distribution function (PDF). Note that there is one such condition for the observed h and k for each individual, as I do not observe commute times for home and commuting method alternatives that individuals did not choose.

B.1.2 Conditional Probability of Observing h and k

I outline the empirical specification of P_i^{HK} for an arbitrary N^H homes and N^K commuting options in individual i's market. $P_i^{HK} = \Pr(U_{ihk} > U_{ih'k'} \,\forall\, (h',k') \neq (h,k) \mid e_{ihk}, \theta)$ is a statement about the joint probability that $N^H N^K - 1$ optimality conditions hold for each individual, conditional on $e_{ihk} = \ln(t_{ihk}) - K_{ihk}\gamma_k^K$. I proceed by first defining the optimality conditions as functions of the errors and the data. Then, I condition on a subset of the errors and recast the optimality conditions in a tractable form for estimation.

Optimality Conditions Observing an individual living in house h and commuting by method k implies that $U_{ihk} > U_{ih'k'} \,\forall\, (h',k') \neq (h,k)$.[64] After algebraic manipulation, it can be shown that these conditions are equivalent to $e_{ih'k'} > f(\xi_i, D_{ih'k'}) \,\forall\, (h',k') \neq (h,k)$, where $f(\cdot)$ is defined to compactly represent the optimality condition as a function of the errors and data; ξ_i was defined in Section 12.1.2 as the vector of unobserved heterogeneity terms; and $D_{ih'k'} = \{c_{ih'k'}, H_{ih'}, K_{ih'k'}\}$ is the set of data that varies over conditions and is used for notational convenience. Explicitly,[65]

$$\begin{aligned} f(\xi_i, D_{ih'k'}) &= \ln\left(T - L_i \vphantom{\frac{1}{1}}\right. \\ &\quad \left. - \frac{T - L_i - \exp\left(\tilde{\lambda}_{i\tilde{k}} + w_{ik}\right) t_{ihk}}{\exp\left(\frac{\ln\left(\frac{c_{ih'k'}}{c_{ihk}}\right) - \exp(X_i\beta^H + \mu_i)\left[\left(H_{ih} - H_{ih'}\right)\gamma^H + \tilde{\varepsilon}_{i\tilde{h}} + \upsilon_{ih} - \tilde{\varepsilon}_{i\tilde{h}'} - \upsilon_{ih'}\right]}{\exp\left(X_i\beta^\ell + u_i\right)}\right)}\right) \\ &\quad - K_{ih'k'}\gamma_{k'}^K - \tilde{\lambda}_{i\tilde{k}'} - w_{ik'}. \end{aligned}$$

[64] Optimality conditions can take one of three forms. Either the individual prefers the observed combination of h and k to

1. Living in another house and commuting by another method ($U_{ihk} > U_{ih'k'} \,\forall\, h' \neq h \,\&\, k' \neq k$),
2. Living in another house and commuting by the observed method ($U_{ihk} > U_{ih'k} \,\forall\, h' \neq h$), or
3. Living in the observed house and commuting by another method ($U_{ihk} > U_{ihk'} \,\forall\, k' \neq k$),

so the notation $(h',k') \neq (h,k)$ is equivalent to $h' \neq h$ and/or $k' \neq k$. Regardless of which of the three conditions is relevant for the given combination of h' and k', I can express the optimality condition as a function of $e_{ih'k'}$.

[65] Note that I replace $\varepsilon_{ih} = \tilde{\varepsilon}_{i\tilde{h}} + \upsilon_{ih}$ and $\lambda_{ik} = \exp\left(\tilde{\lambda}_{i\tilde{k}} + w_{ik}\right)$ when defining this function. Also note that I do not replace t_{ihk} as it is observed directly in the data for the individual's chosen home and commuting method.

Bounds of Integration Evaluation of the joint probability that all of these optimality conditions holds cannot be accomplished analytically or numerically. Instead, I proceed by conditioning on e_{ihk} and ξ_i to make the problem tractable, then evaluating the multidimensional integrals that result using simulation methods. This requires determining the region over which each of the errors in ξ_i are integrated. Placing bounds on the errors being integrated is necessary to avoid situations where draws of the simulated errors are such that no values of the remaining, unintegrated, errors (the $e_{ih'k'}$) are consistent with the data. Figure 2 provides a general, graphical representation of this situation, abstracted to two dimensions. The amorphous, shaded region depicts the values of the

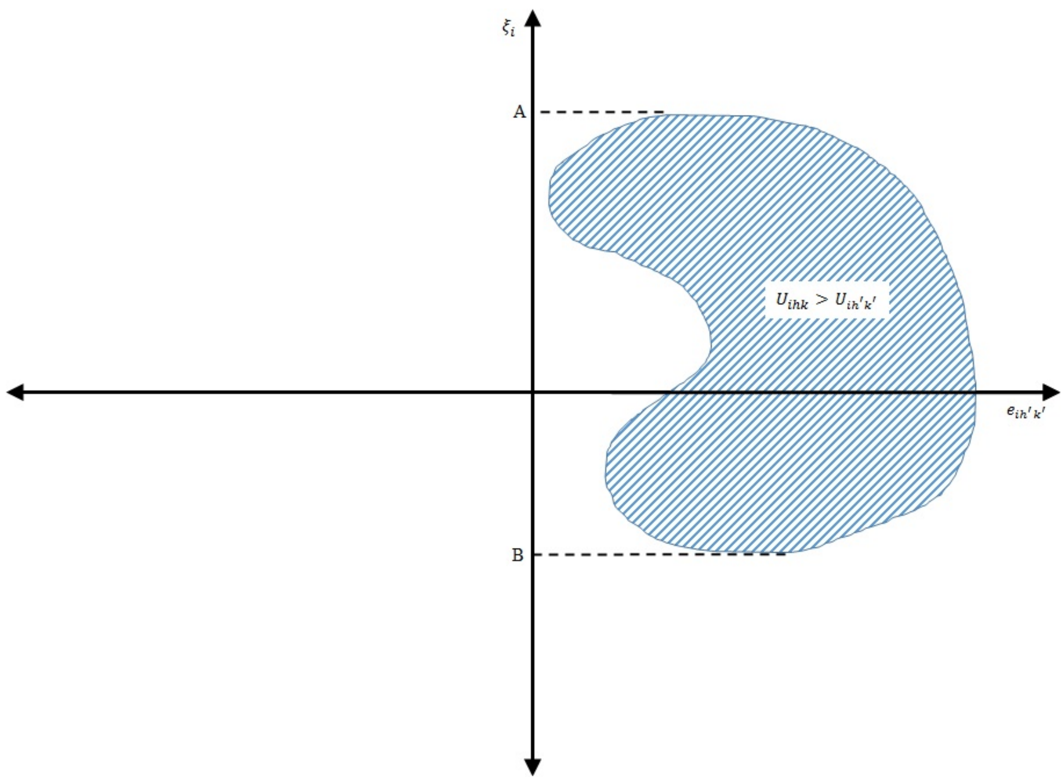

Figure 2: Intuition for Bounds of Integration

errors that are consistent with individual i choosing home h and commuting method k. The errors in ξ_i must be drawn such that $A < \xi_i < B$, otherwise no value of $e_{ih'k'}$ is consistent with what is observed in the data.

Specifically, the value of e_{ihk} is fixed by the estimation of P_i^t (by the relationship that $e_{ihk} = \ln(t_{ihk}) - K_{ihk}\gamma_k^K$), and all of the errors in ξ_i, save for w_{ik}, are integrated over their full distributions. Simulating w_{ik} from an untruncated distribution in this manner would be problematic, as there are some feasible values of w_{ik} for which no value of a given $e_{ih'k'}$ could explain the observed outcomes. This would occur when the random components in the leisure term associated with the observed choices ($\tilde{\ell}_{ihk}$) are such that $\tilde{\ell}_{ihk} \to 0$. Leisure enters the utility function as the argument of a natural log (see Equation 2), so as leisure goes to zero, the utility of the given choices goes to negative infinity. This is not a problem for an unobserved combination of a home and commuting method, as it ensures that $U_{ihk} > U_{ih'k'}$ for the given $(h',k') \neq (h,k)$, however, such a situation is a concern for the observed combination of choices because the model cannot explain an individual choosing options that result in a utility of negative infinity. To ensure that the the optimality conditions can be evaluated, it must be the case that leisure is positive for the observed choices. This necessitates a bound on w_{ik}. It can be shown that the condition that $\tilde{\ell}_{ihk} > 0\, for\, h\, and\, k$ is equivalent to the condition that $w_{ik} < B_i^w$ where

$$B_i^w = \ln(T - L_{\mathrm{i}}) - K_{ihk}\gamma_k^K - \tilde{\lambda}_{i\tilde{k}} - e_{ihk}\, for\, h\, and\, k.$$

Proof that bounding the other errors in ξ_i is unnecessary is straightforward, as the $e_{ih'k'}$ errors enter the optimality condition linearly. Since the support of $e_{ih'k'}$ is the real line, any observed outcome can be justified by an $e_{ih'k'}$ in the appropriate range, so long as $f(\xi_i, D_{ih'k'})$ can be evaluated. The restriction that $w_{ik} < B_i^w$ ensures that the numerator of the second term in the argument of the natural log of $f(\xi_i, D_{ih'k'})$ is positive. The only concern, then, is that the denominator of the second term in the argument of the natural log in $f(\xi_i, D_{ih'k'})$ is such that

$$T - L_i - \frac{T - L_i - \exp\left(\tilde{\lambda}_{i\tilde{k}} + w_{ik}\right) t_{ihk}}{\exp\left(\frac{\ln\left(\frac{c_{ih'k'}}{c_{ihk}}\right) - \exp(X_{\mathrm{i}}\beta^H + \mu_i)\left[\left(H_{ih} - H_{ih'}\right)\gamma^H + \tilde{\varepsilon}_{i\tilde{h}} + \upsilon_{ih} - \tilde{\varepsilon}_{i\tilde{h}'} - \upsilon_{ih'}\right]}{\exp\left(X_{\mathrm{i}}\beta^\ell + u_i\right)}\right)} < 0.$$

This occurs as the denominator goes to zero ($\exp(\cdot) \to 0$). With algebraic manipulation, it can be shown that this is a corner solution. It only occurs when $U_{ihk} > U_{ih'k'}$ for the given $(h',k') \neq (h,k)$ regardless of the value of $e_{ih'k'}$. Intuitively, this means the utility from composite consumption and housing amenities associated with the observed choice is great enough that that it doesn't matter how little time it takes to commute from the unobserved home and/or by the unobserved commuting method, the individual will always choose the observed combination. When this is the case, I do not need to calculate $f(\xi_i, D_{ih'k'})$ to evaluate the probability of the given optimality condition holding.

I define $\bar{\phi}(\xi_i)$ as the joint distribution of ξ_i and B_i as the upper bound on ξ_i. The over-bar on $\bar{\phi}$ denotes that the distribution is truncated for some elements of ξ_i, namely the w_{ik}. Similarly, the bound on the joint distribution of ξ_i is only binding for w_{ik} so

$$B_i = \begin{cases} B_i^w & if\, \xi_{ij} = w_{ik} \\ \infty & else. \end{cases}$$

The probability that ξ_i is less than B_i is equal to the $\Pr(w_{ik} < B_i^w)$. I define this probability as P_i^B where

$$P_i^B = \Phi\left(\frac{B_i^w}{\sigma_w}\right). \tag{6}$$

Joint Probability After integrating over the errors in ξ_i, the optimality conditions can be written in a form that is tractable for estimation. The probability of interest, P_i^{HK}, is the probability that individual i chooses house h and commuting method k. It is expressed as

$$P_i^{HK} = \Pr\left(e_{ih'k'} > f\left(\xi_i, D_{ih'k'}\right) \forall \left(h', k'\right) \neq (h,k) \mid e_{ihk}\right).$$

Using the law of total probability and the assumption that the $e_{ih'k'}$s are independent of both each other and the the other errors in ξ_i, P_i^{HK} can be written as the product of $N^H N^K - 1$ conditional probabilities, so

$$P_i^{HK} = \prod_{(h',k') \neq (h,k)} \Pr\left(e_{ih'k'} > f\left(\xi_i, D_{ih'k'}\right) \mid e_{ihk}, \xi_i < B_i\right) P_i^B.$$

After integrating over the joint distribution of the errors in ξ_i, conditioning on $e_{ihk} = \ln\left(t_{ihk}\right) - K_{ihk}\gamma^K$, and replacing the remaining $e_{ih'k'}$s with their standard normal component according to the relationship $e_{ih'k'} = \sigma_e \eta_{ih'k'}$, I write the joint probability of observing a family living in house h and commuting by method k in a form that is tractable for estimation as

$$P_i^{HK} = \int_{-\infty}^{B_i} \prod_{(h',k') \neq (h,k)} \left[1 - \Phi\left(\frac{f\left(\xi_i, D_{ih'k'}\right)}{\sigma_e}\right)\right] P_i^B \bar{\phi}\left(\xi_i\right) d\xi_i. \tag{7}$$

B.2 Identifying Variation

I encourage the reader to refer back to Equation 2, the individual's full choice problem, and Section 12.1, that details the model's error structure, while reading this section. I begin by discussing the identification of the commute time parameters, which are the coefficients in the commute time equations (γ^K) and the standard deviation of the commute time error (σ_e). The amount of time an individual reports taking to travel from her home to her job depends on the distance between the two locations and the speed the individual travels. The GIS commute characteristics I produce (GIS predicted times) are used to capture the effects of these factors on commute time. The commute characteristic parameters, γ_k^K, are identified by the covariation of commute characteristics (K_{ihk}) with the commute time the individual reports to the ACS (t_{ihk}). There is no guarantee that the individual will choose to travel the exact route mapped by the GIS algorithm, and even after conditioning on route, the characteristics are not perfect descriptors because of congestion, speeding, variation in mass transit schedules, and measurement error in the network data. This means that model-predicted commute times will deviate from the observed times. Variation in these deviations identifies the standard deviation of the commute time error, σ_e.

The error associated with commute time, e_{ihk}, is necessary, but not sufficient, to explain why individuals do not always commute by what the model determines is the optimal method. Individuals choose commutes based on considerations other than financial and time costs. Some individuals in the data choose to commute by a method that is more expensive, both in terms of money and time, than a given alternative. This can be explained by the individual having a high preference for the costlier method, be it because an automotive commuter enjoys listening to music in his car or a mass transit commuter enjoys reading the paper on the subway.[66] The method-specific λ_{ik} error accounts for these preferences. It is separately identified from e_{ihk} by the exclusion restriction that e_{ihk} varies with homes, because of error in predicting commute-location-specific-times, but λ_{ik} does not. The variance-covariance parameters in Ω^{λ} are identified as in other polychotomous discrete choice models (see Bunch, 1991). The intuition for the identification of these parameters is that if an individual does not choose the commuting method that results in the greatest utility according to the model (for the purposes of exposition, say automobile), then the unobserved preference for automotive commuting, $\lambda_{i,auto}$, must be such that it was not the best option ($\lambda_{i,auto}$ is large relative to other λ_{ik}). If when individuals do not select commuting by car, they frequently do not select another given method (say, carpool), then there is a positive correlation between the unobserved preference for those commuting methods. Alternatively, if individuals do frequently select carpool when the (hypothetical) model-predicted best option of automotive commuting is not chosen, then there is a negative correlation between commuting by car and by carpool.

Now I consider the parameters involved in the housing choice, γ^H and Ω^{ε}. Individuals choose a home based on its intrinsic characteristics (e.g. number of rooms), locational characteristics (e.g. proximity to mass transit), and cost. The covariation of observable housing characteristics and the observed housing choice identifies the γ^H parameters. There are assuredly additional characteristics of the home that the econometrician does not observe. An individual may prefer an open floor plan and choose a large home with few rooms. An individual may select a home because it is close to family members (or select a home that is on the other side of town). The housing-specific error term, ε_{ih}, is necessary to explain cases where an individual selects a home that is observationally inferior to other homes in her feasible choice set. The variance-covariance parameters in Ω^{ε} are identified as in other polychotomous discrete choice models (again, see Bunch, 1991). The intuition in this case is similar to the intuition for identifying the parameters in Ω^{λ}. If an individual does not choose the home with the highest observable quality ($H_{ih}\gamma^H$) she can afford, the unobserved preference for that home, ε_{ih}, must be such that it was not the best option (ε_{ih} is negative or relatively small if positive). If when individuals do not select that home, they frequently also do not select another given home, then there is a positive correlation between the unobserved quality of those homes. If, on the other hand, individuals do frequently select the other given home, then there is negative correlation between the errors.

Finally, while the errors mentioned previously are necessary to explain deviations from the predicted optimal housing and commuting methods separately, the joint decision of housing and commuting method needs to be explained as well. The random preference parameters, α_i^c, α_i^H, and α_i^{ℓ}, are necessary to explain deviations from the predicted joint decision. As I am modeling a discrete choice, I must normalize one of the parameters, as the level and scale of utility are

[66] Note that it can also be explained by the individual having a low preference for the unchosen option because she finds driving on congested roads to be stressful or because she does not like to stuff herself into a crowded bus.

irrelevant. I do so by setting $\alpha_i^c = 1$, which addresses the issue and is equivalent to fixing one of the variance terms.[67] The remaining parameters account for the fact that even if two individuals value all homes and commutes the same, they may be observed living in two distinctly different homes and commuting by different methods. This would occur if they had different relative preferences for composite consumption, housing amenities, and leisure time. I provide intuition with three illustrative examples. In all, I assume that two individuals agree in their valuations of housing and commuting options, and they both commute to the same location.

1. Assume that these two individuals live in homes that are identical in every way, save location. The first lives in a home that is closer to their shared job location, so he has a shorter commute, but that commute is more financially costly than the commute taken by the second individual. The former has a greater preference for leisure relative to composite consumption than the later, so $\alpha_1^\ell > \alpha_2^\ell$.

2. Now assume that the two individuals are neighbors as well as coworkers, so they have identical commutes, both in terms of time and financial costs. If the first individual lives in a better, more expensive home than the second, then he prefers consumption of housing amenities to composite consumption, so $\alpha_1^H > \alpha_2^H$.

3. Finally, assume that these two individuals live in homes that are of equal cost and commute by methods of identical financial cost. The first lives in a downtown apartment that is close to their shared job location. The second lives in a suburban home that is farther from work, but has more housing amenities than the downtown apartment. The former has a greater preference for leisure relative to housing consumption than the later. This indicates that $\frac{\alpha_1^\ell}{\alpha_1^H} > \frac{\alpha_2^\ell}{\alpha_2^H}$.

Regardless of case, the covariation of the observable individual characteristics (X_i) and consumption of housing ($\tilde{H}_{ih}$) with housing and commuting outcomes identifies the β^H parameters. Similarly, the covariation of the observable individual characteristics (X_i) and leisure ($\tilde{\ell}_{ihk}$) with outcomes identifies the β^ℓ parameters. The individual observables will not perfectly predict the preference parameters, hence the inclusion of error terms associated with the individual's preference for housing amenities (μ_i) and leisure (u_i) in the model. Correlation between higher order moments of the deviations and higher order moments of the consumption of housing and leisure identifies the variance parameters in $\Omega^{\vec{\mu}}$.

B.3 Threats to Identification

Although this work advances the literature in several important ways, assumptions are necessary to keep the model tractable. As stated earlier, I assume that an individual takes her city of residence, family structure, vehicle ownership, and employment as given; and that the locations and hours of

[67] See Train (2009) for an excellent treatment of the subject.

firms and schools are independent of residential choices and transportation options. I discuss the potential bias that each of these assumptions introduces in the remainder of this section.

Defining the local residential market as closed at the metropolitan level is necessary to limit an individual's choice set when searching for a home. It has the potential to bias results to the degree that individuals select their city of residence based on characteristics of the residential or commuting markets in the city. For instance, if an individual chose to locate in a city because of a lack of congestion or the availability of a particular commuting option, my model would understate the preference that the individual has for those amenities. On the other hand, if employment opportunities alone drive the choice of city, then this source of bias might arise only if firms choose locations on the same basis, which I will not be able to model.

Excluding family structure decisions, such as fertility, are another possible source of bias. For example, Dettling and Kearney (2011) find that changes in house prices have differential effects on the birth rates of home owners and non-owners. If an individual decides to have children because their home location is more conducive to raising children, my model will overstate the impact of those children on the individual's value of the given housing amenities. A similar logic applies to the sign of the bias that children might cause on commuting amenities (e.g. a shorter or more flexible commute). These concerns are an interesting topic for future research.

Ignoring automobile ownership decisions is a more problematic assumption in the direct context of my model. My model removes commuting by car from the choice set of a household that does not own an automobile, but an individual who does not own a car may do so because she has a high distaste for commuting by car. My model will understate this individual's distaste for commuting by car, but explicitly modeling automotive ownership decisions is not supported by the available data. I observe very little about automobile ownership: only how many vehicles are available for use by members of the household. Fortunately, concern over this bias is mitigated by the fact that the automobile ownership rate is quite high: 87.4% of the families in my sample have at least one car per adult in the family.

Assuming that labor market decisions are exogenous also is not benign. In my model, I treat individuals as searching for a place to live subsequent to finding a job. However, the converse could also be true. This causes bias if, for example, an individual with a high distaste for commuting trades proximity to her home for wages when accepting employment. If so, my model would return a biased estimate of this individual's aversion to commuting, as it will explain some of the residential choice as a function of low wages preventing the individual from being able to afford a long commute, understating the individual's distaste for commuting. It is important to note that all of the residential choice studies I cite in Section 2 make a similar assumption. The alternative would require modeling job search behavior, which is not possible given the available data, as I observe only minimal characteristics of the individual's current job.

The assumption that the locations and hours of firms and schools are independent of my choice variables is implausible. Both are likely to locate in response to the distribution of residential housing and factor local commuting conditions into their decision of how to set their hours of operation. Again, I provide an example of how this might lead to bias. If firms locate close to neighborhoods where a critical mass of individuals reside, my model will overstate the aversion those workers have to commuting long distances. I justify this assumption similar to the justification for the assumption that the agents in a perfectly competitive market are price takers by assuming that any individual's choice of residence and method of commuting can neither influence where nor when firms and schools operate.

Despite these shortcomings, it is important to remember that that my model makes several key advances by jointly modeling residential choice and commuting method at the individual level in a way that allows for a rich heterogeneity structure and incorporates collective household decisions. I remind the reader that Langer and Winston (2008) propose a joint model of residential choice and commuting mode similar to the one I outline but opt for a different research design because "the data and modeling requirements of a disaggregate approach... are formidable."