Choice of residence location and mode of commuting: a cross-sectional analysis of 275 US metropolitan areas*

Debarshi Indra**

7/15/2014

ABSTRACT

A nested logit model is employed to capture workers' joint choice of residence location and mode of commuting in a U.S. metropolitan area. The nested logit model is estimated using tract level aggregated data from the U.S. Census Bureau's transportation planning package (CTPP) and summary file 3 from the 2000 U.S. decennial census. The estimation covers worker flows of more than four million work-residence census tract pairs contained in 275 metropolitan areas. The effects of accessibility to water bodies, limited access highways, central cities, and consumption opportunities in workers' decision process are considered. The nested logit model is estimated using a national sample pooled across all MSAs and individual MSAs. We find that the mode choice elasticity with respect to commuting time has declined in the U.S. and that it follows the Burr distribution in the population of MSAs. Higher mode choice elasticity is also found to increase MSA population density or decrease urban sprawl.

Keywords: Discrete choice, urban economics, location choice, mode choice

JEL classification: C35, R14, R15, R21, R41

**The author is a Post-Doctoral fellow at the Edward J. Blakely Center for

Sustainable Suburban Development, University of California Riverside.

indradebarshi@gmail.com.

***Acknowledgment:** The author acknowledges the support of the Multi-campus Research Program and Initiative (MRPI) grant from the Office of the President, University of California (award number 142934). Views expressed in the article are solely those of the author and not of the financial supporters. The author is grateful to Alex Anas , Richard Arnott and Neel Rao for valuable comments and suggestions. However, the author is solely responsible for any errors.

1. Introduction

We use discrete choice theory to model workers' joint choice of residence location and mode of commuting in a U.S. Metropolitan Statistical Area[1] (MSA). The discrete choice model involves workers' selecting a work - residence census tract pair that falls within the limits of the MSA, and choosing from four modes of commuting. We estimate the discrete choice model using census tract level aggregated data by combining the Census Transportation Planning Package (CTPP) and the U.S. census Summary file 3 datasets from the 2000 U.S. decennial census covering 275 MSAs.

In urban economics, the most popular theoretical framework is the monocentric city model. The standard monocentric model is characterized by continuous space, centralized employment and a single mode of commuting. However, as documented by Mieszkowski and Mills (1993), U.S. cities have been experiencing job decentralization for many decades. In 2000, only 47.73% (Table 1) of MSA jobs were located in U.S. central cities. Most European cities have also experienced job decentralization in the last many decades. With increasing highway construction, rising demand for land and automobiles, owing to higher household incomes, job decentralization can be expected to accelerate in emerging economies as well[2].

The single mode assumption in the standard monocentric city model might be reasonable for most cities in the U.S., where the automobile is by far the most popular mode of commuting (Table 1). However, the single mode assumption certainly does not hold true for many cities in

[1] We use year 2000 Metropolitan Statistical Area definition.

[2] All these factors are believed to have contributed to job decentralization in the U.S.

other parts of the world. For example, in Europe and most emerging nations, public transportation is still a widely used mode for commuting.

The above discussion implies the need for an alternative to the monocentric city model which is capable of incorporating multiple employment centers and modes of commuting. However, the reason for the popularity of the monocentric city model among urban economist in the face of contrary evidence is that the analytical difficulties in extending the monocentric city model to include multiple employment centers and modes of commuting are considerable (see, for example, White 1988 and LeRoy and Sonstelie 1981). The analytical complexities of these models also make them ill-suited for estimation with real world data where multiple employment centers and modes of commuting are the norm.

An alternative approach to construct a tractable model of location choice and land use to include multiple employment centers and mode of commuting is to use discrete choice theory (Anas and Liu 2007). Discrete choice based models have an advantage from an empirical point of view because they incorporate taste heterogeneity in an attractive manner. In discrete choice models, the utility functions of workers' contain idiosyncratic effects for each discrete alternative which are randomly distributed. This means that the demand functions are expressed as probabilities obtained by integrating over the distribution of the idiosyncratic effects, and can be taken straight to data for estimation. This is unlike continuous models, where taste heterogeneity is included by adding error terms to the demand function before estimation. In this sense, land use models based on discrete choice theory provide a stronger link between theoretical and empirical work.

Discrete choice models have a long history in the empirical urban housing market literature as well, and have been used by economists to study tenure, residential location and dwelling unit choices, (Quigley 1984, Rapaport 1996, Yates and Mackay 2006, and Ioannides and Zabel 2007) and in the transportation field where commuting mode choice is usually the main focus (Small and Verhoef 2007).

We note three shortcomings about the empirical residential location choice literature some of which follow from the weaknesses of the monocentric city model mentioned above. First, most empirical studies on residential location choice ignore commuting mode choice. However, as Glaeser et.al (2008) find, the availability of public transportation might be a strong reason for the concentration of poor households in some U.S. central cities. Therefore, commuting mode might influence residential location, for at least some households, and should be included in analyzing residential location choice. Second, most studies ignore the effects of workers' work location on their residence location choice, which again may reflect the popularity of the monocentric city model in the urban economics literature. If all jobs are centralized in one location, one can safely ignore the effects of workers' work location as they would not vary across workers. However, as White (1988) shows, in a city with decentralized employment, identical workers will choose different residential locations depending on their work location. Wadell et.al (2007) accounts for the effects of work location on household location choice by analyzing a joint choice model of work and residence location for the Puget Sound region of Washington State. Third, most studies of residential location choice are restricted to one city or area. An exception is Ioannides and Zabel (2008), who study residential location choice for about 100 U.S. metropolitan areas using micro level data from the American Housing Survey and the U.S census bureau. However, Ioannides and Zabel (2008) do not consider commuting mode choice, or control for the workers' work location. Bento et.al (2005) study the effects of urban structure on mode choices and VMT in 114 U.S urban areas, however, they ignore residential location choice.

This paper attempts to address the points raised above for the empirical residential location choice literature by extending the work of Anas and Chu (1984). They estimate a joint choice model of residential location and mode of commuting using aggregated data for the Chicago MSA. However, Anas and Chu (1984) consider only workers commuting to the central business district in downtown Chicago from some randomly selected small zones which were formed by aggregating adjacent census tracts. We do not aggregate census tracts, and include all work - residence census tract pairs which are linked by commuting flows that lie within the limits of 275 MSAs' in the U.S. Like Anas and Chu (1984) we consider four different types of commuting modes. We also examine the effect of workers' job location on residential location choice by including the central city or suburban status of the workers' work location.

Following Anas and Chu (1984) we estimate the model using data that the U.S. Census has aggregated to the census tracts. The reason for using aggregated data from the U.S. census bureau is that it is geo-referenced. We know to a high degree of precision workers' work – residence location pair within a MSA. The location information is suppressed in most publicly available national micro data sets to protect the respondents' privacy. For example, the National Household Travel Survey (NHTS), which collects data on workers' travel patterns and travel modes in the U.S., does not reveal workers' work or residence location within a MSA. The NHTS also does not provide the MSA location for workers living in small MSAs. The same location suppression is done by the American Housing Survey (AHS) which collects data on individual residential houses in the U.S. The location information is particularly important if we want to control for the geographical neighborhood of the worker. It is also essential to compute accessibility measures. In addition, unlike the NHTS and the AHS, the aggregated data from the 2000 U.S. census contains information on both neighborhood housing characteristics and workers' commuting. Finally, the 2000 U.S. census had an average nationwide sampling rate of 1-in-6 households implying a comprehensive national coverage. When projected to the population, the CTPP 2000 therefore consists of 4,372,582 work-residence census tract pairs, covering the commuting decision of 95,854,466 workers nationwide.

An important issue that one has to deal with in the discrete choice based residential location choice literature is the creation of the workers' choice sets. If the study area is divided into a large number of locations, for example census tracts, compiling the choice set for each worker in the sample to contain every location might make estimation computationally infeasible. For the multinomial logit model, a consistent estimator can still be obtained if one performs the estimation on random samples of the workers' choice sets (Train 2009). This result does not extend to the nested logit model which we use in this paper (Nerella and Bhat 2004). Given the size and nature of the data set, we therefore employ a simple procedure to create the workers choice sets used in estimation. The choice set for each worker, given their work location, includes all residence location and modes of commuting combinations in the MSA which are available from the worker's work location and are observed to be the actual choices of some commuters in the data.

We estimate a nested logit model using a national sample which pools data across all MSAs. We also estimate models for different income groups from the national sample. Finally, using the national nested logit model as a benchmark we estimate nested logit models for each of the 275 MSAs in our sample. For these selected MSAs we pool workers across all household income groups.

Our results are mostly in line with expectations. We find from the nationwide estimations that regardless of workers' income backgrounds, the most preferred alternative is working and living in the suburbs and driving alone to work. We also find that workers' belonging to low income households prefer not to locate in high income census tracts. This evidence points to some kind of prejudice against the poor. Our estimates of mode choice elasticity with respect to commuting time, which compare well with recent micro data based studies, shows that the mode choice elasticity in the U.S. has come down as compared to previous decades.

The MSA specific estimations yield distributions for two elasticities of interest: mode choice with respect to commuting time and residential location choice with respect to housing costs. We find that the distributions of these elasticities in the population of MSAs follow closely the Burr (Singh-Maddala). As these elasticities describe workers preferences, and MSA population density is an outcome of workers' choices based on these preferences, there should be some relationship between the two elasticities and MSA population density. A regression of population density on these two elasticities reveal that, increasing mode choice elasticity increases population density while increasing housing cost elasticity lowers population density. To check the robustness of our result we run regression of a MSA sprawl index (Burchfield et al. 2006) on these elasticities and find just the opposite effects. Since MSAs with more transit infrastructure have higher mode choice elasticity, an implication of these results is that more public transportation might raise MSA population density or lower sprawl.

The paper proceeds as follows. Section 2 describes the data set in more detail. Section 3 describes the nested logit model and section 4 describes the estimation procedure. In section 5 we discuss the estimation results, and in section 6, the elasticities derived from the estimates. We conclude in section 7 by pointing to some improvements, extensions and possible applications of the results.

2. Data

This paper uses three kinds of variables, which describe workers' work location, residence location, and work-residence location pair. Most of these variables come from the 2000 U.S. decennial census, aggregated to the level of census tracts. The 2000 Census Transportation Planning Package (CTPP), parts1-3 provides data by workers' residence location, work location, and work-residence location pair and mode choice, respectively. However, the CTPP contains limited data on housing characteristics in a census tract. The housing data for each census tract comes from the 2000 U.S. census summary file 3. We also use information from the National Atlas of the U.S. to construct some geographic attributes of a census tract. We now briefly describe the variables mentioned in Table 2.

We use the CTPP part three to find the flow of workers between work - residence census tract pairs by mode of commuting and household income. The modes of commuting include drove alone, carpooling, public transportation and other modes[3]. The household income groups are, less than $30,000, $30,000-$49,999, $50,000-$74,999, and $75,000 and above. CTPP part 3 also gives the average commuting time between work-residence census tract pairs by mode of commuting. Our data set contains the flow of workers (by mode of commuting and household

[3] Other modes include bicycle, walked, taxicab, motorcycle or other means, and worked at home.

income) and average commuting times (by mode) for all work-residence census tract pairs located within 275 metropolitan areas in the U.S. We restrict our analysis to workers who work and reside in the same metropolitan area ignoring the minor flow of workers who live and work in different but adjacent metropolitan areas. Our dataset covers the flow of around 95 million workers commuting between approximately 4.3 million census tract pairs in year 2000.

The Census Bureau, due to privacy concerns, censors the mode of commuting and household income category for around 30 million workers, out of a total of 95 million workers. The censoring takes place when the total count of workers commuting between a pair of work-residence census tract falls below a certain threshold set by the Census Bureau. As the Census Bureau employed a variable sampling rate for census 2000, the threshold is not the same for all census tract pairs. In most cases a count of four workers will fail to pass the threshold. However, in some cases, a count of as many as 20 workers might also fail to pass the threshold. The Census Bureau, however, does provide data on aggregate and average commuting times by mode of commuting for all work-residence census tract pairs, which includes those subject to censoring. From the data on aggregate and average commuting times, we impute the flow of workers by mode of commuting (by dividing the aggregate travel time by the mean travel time) for those work-residence census tract pairs subjected to censoring. We, however, cannot impute the household income category for those workers whose work-residence census tract pairs were subjected to censoring. Therefore, these imputed commuting count data will be used in our models in which all income groups are pooled but not in the income group specific models.

Using CTPP part 2 we construct approximate worker earnings by household income group. The earnings are computed as the weighted average of the midpoints of earning categories, the weights being the number of workers who belong in that category. This method leads to earnings which are usually biased downward, due to earnings being top coded at \$75,000. An alternative method would involve fitting some distribution to the earnings data. However, this exercise is difficult to implement with 50,000 workplace census tracts, and each tract potentially having a different earnings distribution.

We also use CTPP part 2 to construct a census tract's accessibility to consumption opportunities. The formula for the accessibility measure is shown below and measures the accessibility, A_j, of a residential tract *j* to consumption opportunities at tracts $k = 1, \ldots, N_j$ tracts accessible from *j*:

$$A_j = \sum_{k=1}^{N_j} jobs_k \exp\left(-\tau w_{jk}\right) \tag{1}$$

$jobs_k$ is the proportion of MSA retail/arts/entertainment jobs, per square mile, of census tract k. w_{jk} is a travel impedance weight attached to census tract k. τ is a parameter that we set[4]. We have two choices for the weight. One choice is the straight line distance between census tracts *j* and *k*. The advantage of using distance as weight is that it is available for all work-residence census tract pairs in the sample. In this case N_j will be same for all *j*, and the accessibility measure will include all census tracts in a MSA. The other, and in our opinion, a better candidate for w_{jk} is the travel time between census tracts *j* and *k*, for non-commuting travel. The reason would be that much consumption-related travel from homes avoids the peak-period during which most commutes occur. However, as census data does not contain non-commuting travel times, we have to use average commuting times from CTPP part 3 to proxy for

[4] We experimented with a number of values for τ before settling for $\tau = 0.01$ as it gives enough variation in the data to make estimation meaningful.

non-commuting travel times. But, using commuting times means that in the accessibility measure we have to drop census tract pairs which have zero workers commuting between them, and for which –therefore – commuting time is not observed. Then, N_j will include only those census tracts which receive commuting flows from census tract *j*. This turns out to be reasonable, since given our very large sample, if no worker in a metropolitan area is observed commuting between a pair of census tracts; this implies that travel between that pair of census tracts might be difficult or infeasible, for non-commuting purposes as well.

We use CTPP part 1 to compute the average number of workers per household for each household income category. We divide housing cost per household by the average number of workers per household to get the average housing cost per worker.

Each census tract in the data set comes with the latitude and longitude of its internal point[5]. We use it to compute the straight line distance to work. We also use the internal point to determine if a census tract is located in a MSA central city. We define a census tract to belong to a MSA central city if its internal point falls within a central city[6]. Using this definition, we find census tracts in MSAs which belong to their respective central cities. A detailed report on the GIS methods used in this paper is available from the author upon request. The above procedure is then used to create two dummy variables, one for work census tracts and the other for residence census tracts. These dummy variables take the value of one if a work or a residence census tract belongs to a central city, and zero otherwise.

---

[5] The internal point of a census tract is the geographic center of the polygon that defines the boundary perimeter of the.

[6] We used central city boundary map layer from the U.S Census Bureau.

The National Atlas of the U.S provides a variety of map layers to the public free of charge. The hydrography map layer includes the coastline, major rivers, streams and canals; and major lakes and reservoirs of North America. The road map layer contains an extensive network of roads in the U.S. We use the hydrography and road map layers to construct a map layer which contains the U.S. coastline, some major rivers, lakes and the entire network of limited access highways. This map layer is shown in figure 1. This map layer is then used to construct two accessibility variables, distance to water and distance to road[7].

The variables median age of house, median number of rooms in a house, and land area of a census tract attempt to characterize the housing characteristics of a census tract. The total number of housing units is used to measure housing stock in a census tract.

Equation 2 shows the formula used to compute annual housing cost.

$$\text{Annual housing cost } = 0.1 \times V \times S + mcr \times (1 - S) \tag{2}$$

V is the median value of all owner occupied housing units, *mcr* is the annual median contract rent for renters, and *S* is the share of occupied housing units that are owner occupied. 0.1, informally known as Shelton's rule in real estate studies, is used to convert housing value to annualized user cost (or imputed rent) for owner occupied housing units (Anas, 1981). Note that, the variable annual housing cost captures the housing cost of a household. We convert household cost to worker cost by dividing annual housing cost by the average number of workers in a household, which-as explained earlier- we get from CTPP part 1.

---

[7] Distance refers to the straight line Euclidian distance between the internal point of a census tract and the nearest map feature.

Median household income in a census tract attempts to proxy for public infrastructure in a census tract. For example, census tracts with higher average household income are more likely to have better schools, better parks, etc. This especially matters for workers with children who are choosing their residence location.

In Tables 3a, 3b, and 3c we present the descriptive statistics by the census division of the country, for all the variables discussed above.

3. Model

In a discrete choice model an agent chooses one alternative from a set of mutually exclusive alternatives. In this paper, the agent is a worker, and an alternative consists of a residence location and a mode of commuting, given the worker's work location in a MSA. Let i index the worker's work location, j the residence location and m the mode of commuting. As mentioned earlier, the worker's choice set includes all residence locations and modes of commuting in the MSA which are observed in the data given the worker's work location.

$$U_{jm|i} = \beta Y_{j|i} + \delta Z_{m|ij} + \varepsilon_{jm|i} \tag{3}$$

$U_{jm|i}$ denotes the worker's utility associated with a choice alternative (j,m). $Y_{j|i}$ is a vector of attributes that vary by residence location j given workplace location i, and $Z_{m|ij}$ is a vector of attributes that vary by mode of commuting, given residence location j and workplace location i. $Y_{j|i}$ and $Z_{m|ij}$ include observable attributes of an alternative. β and δ are vectors of parameters that we will estimate from data.

$\varepsilon_{jm|i}$ is the component of worker's utility from the alternative which is unobservable to the researcher. $\varepsilon_{jm|i}$ is therefore modeled as random utility, that is, a random variable for each worker in the population of workers. The presence of this random component in utility implies that the worker's utility $U_{jm|i}$ is a random variable and so we can only make probabilistic statements about the worker's potential choices.

We assume that the vector of random utilities $\langle \varepsilon_{jm|i} \rangle$ is distributed GEV (generalized extreme value), such that a utility maximizing worker's probability of choosing an alternative is a nested logit (Train, 2009). The nested logit model implies a particular form of nesting and correlation among the worker's random utility components, and can be derived from the GEV model. See Figure 2 for the nesting structure assumed in this paper. According to this nesting structure, $\varepsilon_{jm|i}$ which belong to alternatives that have the same work-residence location pair can be correlated; but the model assumes that there is no correlation among $\varepsilon_{jm|i}$ which belong to alternatives with different work-residence location pairs.

An advantage of the nested logit model is that the joint choice probability can be decomposed into two logit probabilities and can be written as follows:

$$P_{jm|i} = P_{j|i} P_{m|ij} \tag{4}$$

where $P_{jm|i}$ is the joint probability that the worker chooses an alternative. $P_{jm|i}$ can be written as the product of $P_{m|ij}$, the worker's probability of choosing mode m, conditional on his work and residence locations (i,j), and $P_{j|i}$, the worker's probability of choosing residence location j, conditional on his work location i. $P_{m|ij}$ and $P_{j|i}$ are each logit:

$$P_{m|ij} = \frac{\exp\left(\delta Z_{m|ij}\right)}{\sum_{m'=1}^{M_{ij}} \exp\left(\delta Z_{m'|ij}\right)} \tag{5}$$

$$P_{j|i} = \frac{\exp\left(\lambda\sigma_{j|i} + \beta Y_{j|i}\right)}{\sum_{j'=1}^{J_i} \exp\left(\lambda\sigma_{j'|i} + \beta Y_{j'|i}\right)} \tag{6}$$

$$\sigma_{j|i} = \log\left(\sum_{m=1}^{M_{ij}} \exp\left(\delta Z_{m|ij}\right)\right) \tag{7}$$

where M_{ij} are the sets of the commuting modes available to workers working at *i* and residing in *j*; and J_i are the sets of residence locations available to workers, working at *i*. $\sigma_{j|i}$ is known as the "inclusive value" or "log sum" term. These inclusive values measure the maximum expected utility from mode choice, given any work residence location pair (*i*,*j*). The inclusive value connects the two logit probabilities by entering as an attribute in the residence location choice probability.

The coefficient of the inclusive value λ is a measure of the degree of correlation between unobserved utilities. λ equal to one implies that there is no correlation between the unobserved utilities of the travel modes, and the joint choice probability $P_{jm|i}$ becomes a multinomial logit. λ equal to zero implies perfect correlation of the travel mode specific random utilities. If λ is greater than one then the nested logit model is still consistent with utility maximization under certain conditions. But if λ is less than zero, then the nested logit structure is inconsistent with utility maximization (Train 2009).

Note that writing the joint choice probability in terms of two sequential logit probabilities need not imply any kind of sequence in the worker's decision process. It simply reflects the particular structure of correlation among unobserved utility in a simultaneous choice process.

But the decomposition is also helpful in the estimation stage, as we will discuss in the next section..

4. Estimation

In our dataset we do not observe the vector of attributes $\langle Y_{j|i,s}, Z_{m|ij,s}\rangle$ faced by a particular worker, denoted by s hereafter. We observe the averages of the attributes $\langle \bar{Y}_{j|i}, \bar{Z}_{m|ij}\rangle$. As shown below, we assume that the actual values of the attributes faced by the worker are pure random deviations from their averages, where the vector $\langle \xi_{j|i,s}, \zeta_{m|ij,s}\rangle$ represents pure white noise.

$$Y_{j|i,s} = \bar{Y}_{j|i} + \xi_{j|i,s} \tag{8}$$

$$Z_{m|ij,s} = \bar{Z}_{m|ij} + \zeta_{m|ij,s} \tag{9}$$

In the utility function, we substitute for the actual attribute values by their average values plus the deviation from the average. As $\langle \xi_{j|i,s}, \zeta_{m|ij,s}\rangle$ are unobservable to the researcher, they are absorbed into the random component of utility $\varepsilon_{jm|i,s}$ which we now label as $\varepsilon^*_{jm|i,s}$. We still assume that the vector $\varepsilon^*_{jm|i,s}$ is distributed GEV such that the joint choice probability of residence location and mode of commuting is given by the nested logit model. However, now the joint choice probability is identical across workers having the same work location.

$$U_{jm|i,s} = \beta\left(\bar{Y}_{j|i} + \xi_{j|i,s}\right) + \delta\left(\bar{Z}_{m|ij} + \zeta_{m|ij,s}\right) + \varepsilon_{jm|i,s} \tag{10}$$

$$U_{jm|i,s} = \beta\bar{Y}_{j|i} + \delta\bar{Z}_{m|ij} + \beta\xi_{j|i,s} + \delta\zeta_{m|ij,s} + \varepsilon_{jm|i,s} \tag{11}$$

$$U_{jm|i,s} = \beta\bar{Y}_{j|i} + \delta\bar{Z}_{m|ij} + \varepsilon^*_{jm|i,s} \tag{12}$$

$$\varepsilon^*_{jm|i,s} \equiv \beta\xi_{j|i,s} + \delta\zeta_{m|ij,s} + \varepsilon_{jm|i,s} \tag{13}$$

$$P_{jm|i,s} = P_{m|ij}P_{j|i} \tag{14}$$

$$P_{m|ij} = \frac{\exp\left(\delta\bar{Z}_{m|ij}\right)}{\sum_{m'=1}^{M_{ij}}\exp\left(\delta\bar{Z}_{m'|ij}\right)} \tag{15}$$

$$P_{j|i} = \frac{\exp\left(\lambda\sigma_{j|i} + \beta\bar{Y}_{j|i}\right)}{\sum_{j'=1}^{J_i}\exp\left(\lambda\sigma_{j'|i} + \beta\bar{Y}_{j'|i}\right)} \tag{16}$$

We estimate the parameter vector $\langle\beta, \delta\rangle$ by using the maximum likelihood procedure. The likelihood function is shown below, where $N_{jm|i}$ is the number of workers in the data set who choose alternative (j,m) given i. This form of the likelihood function follows from the fact that probabilities are identical across workers with the same work location, as explained above.

$$l = \prod_{i=1}^{I}\prod_{j=1}^{J_i}\prod_{m=1}^{M_{ij}} P_{jm|i}^{N_{jm|i}} \tag{17}$$

The nested logit model allows us to decompose the likelihood function into two parts, one part corresponding to the mode choice model, the other part corresponding to the residence location choice model. This decomposition is shown below for the log of the likelihood function. $N_{j|i}$ and $N_{m|ij}$ are the number of workers who choose residence location j given i, and mode m given i, j, respectively.

$$ll = \sum_{i=1}^{I}\sum_{j=1}^{J_i} N_{j|i}\log\left(P_{j|i}\right) + \sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{m=1}^{M_{ij}} N_{m|ij}\log\left(P_{m|ij}\right) \tag{18}$$

Instead of maximizing the log likelihood function ll with respect to the parameters $\langle \beta, \delta \rangle$, we proceed in a two-step fashion. We first maximize $\sum_i^I \sum_j^{Ji} \sum_m^{Mij} N_{m|ij} log(P_{m|ij})$ with respect to δ and compute the inclusive value for each work residence location pair. We then maximize $\sum_i^I \sum_j^{Ji} N_{j|i} log(P_{j|i})$ with respect to β.

This two-step procedure yields consistent, but less efficient estimators as compared to maximizing ll with respect to all the parameters (Train, 2009). The two-step procedure, however, offers two advantages as compared to the one-step procedure. First, the log likelihood function ll is not well behaved and might not have a unique maximum. This means that numerical maximization procedures might not readily converge to an optimum. However, the log likelihood functions corresponding to the individual logit probabilities are well behaved, and do have a unique maximum. Any numerical maximization procedure will converge smoothly to the respective optima. Secondly, given the size of the dataset, and the number of variables in the utility function; sequential estimation demands less computing resources. Given our large sample size the loss in efficiency from the two step process is not a serious concern. One consequence of the two step procedure is that the standard errors computed in the second step need to be adjusted (Train, 2009). However, since our sample size is large, we ignore that suggestion and do not adjust the standard errors.

Before we conclude this section we allude to two issues which we have avoided so far in our discussion. We did not consider the possibility of spatial correlation and endogenous variables affecting our results. We believe that by including various accessibility and other variables in the utility function we mitigated the problems which might arise from the presence of spatial correlation and endogenous variables. The accessibility measures allow for spatial correlation to be captured through the observed variables and reduce the residual spatial correlation that might remain among the unobserved variables. This allows for our estimators to retain all their desirable properties (Woolridge 2002).

The housing costs in a census tract might be influenced by the census tract's relative location or neighborhood in a metropolitan area, thereby making housing costs a potential endogenous variable. The situation is analogous to the price of a car being affected by certain unobserved qualities of the car. The failure to account for the car's quality then makes the car's price an endogenous variable, since the car's quality is an important factor in the buyer's choice process. In the discrete choice literature many methods have been suggested to solve the potential endogeneity problem (Train 2009). Thankfully, in the case of a census tract, its quality can be captured easily by including various accessibility and other features of the census tract and its housing. Therefore, by including these other variables in the utility function we hope to have remedied any potential endogeneity problems.

5. Estimates

Table 4 presents estimation results for workers' belonging to different household income groups, and results from the pooled sample. The pooled sample is created by including all work-residence census tract pairs that fall within MSAs in the U.S. connected by commuting flows. For the pooled sample, we constrain the parameters to be the same across all metropolitan areas.

Table 4 shows that, nationally, across all income groups, driving alone is the most popular mode of commuting, followed by other modes and public transportation. Surprisingly, carpooling is the least popular mode in the U.S. The dislike for carpooling might be explained by the fact that workers care about privacy, which is least when one is carpooling, or that the time cost involved in picking up passengers at different addresses makes carpooling relatively unattractive. While “other modes” is more popular than carpooling, since it consists of modes like walking, bicycling to work, it is usually feasible only over short distances. As expected, increasing average commuting time, after controlling for mode, lowers utility across all household income groups.

The inclusive value parameter is positive and lies between zero and one for all income categories. Recall, that the inclusive value parameter has important implications in the nested logit model. The fact that the inclusive value parameter always lies between zero and one suggests that the nested logit model is consistent with utility maximization. We find that, as household income rises, the inclusive value parameter falls somewhat, which implies that the degree of correlation across alternatives with the same work-residence location pair increases with income.

One of the main reasons for the existence of cities is that they present their inhabitants with the opportunity to consume a wide variety of goods and services (Glaeser et.al, 2001). However, consumption and travel are intimately related (Anas, 2007). People need to travel to buy goods and services, and such travel incurs both time and monetary costs. Therefore, it is likely that in choosing residence location, workers take into account accessibility to consumption opportunities in their neighborhood. As discussed in the data section (equation 1), we construct a measure of a census tract’s accessibility to retail-arts-and-entertainment opportunities. We include this measure of accessibility to consumption opportunities, and the square of this measure in the utility function to allow for both the positive and negative effects of accessibility. We find a stable U-shaped relationship between utility and accessibility to consumption opportunities across all income categories. This implies that on average the negative effects outweigh the positive effects of accessibility to consumption for resident workers. Workers want to be near such opportunities, but not too near, probably due to the fact that they tend to avoid the hustle, bustle and congestion associated with commercial centers.

Many U.S. metropolitan areas began life near major water bodies (Anas et.al 1998) because of transportation reasons before the era of automobiles and highways. A census tract's distance to water bodies like oceans, lakes and rivers, therefore, may reflect important historical effects. In addition, locating one's residence near a water body has advantages, like proximity to a natural amenity, and disadvantages, like effects of various kinds of negative externalities like blight, and visual pollution and proximity to effluent discharging production that is often associated with certain water bodies, especially rivers, or congestion and hustle and bustle especially if the census tract is near a major tourist destination or a harbor. To capture both positive and negative effects, we include both distance to water and the square of distance to water in the utility function. We assume that the distance to water affects utility, only if, the metropolitan area is located a certain distance to a major water body: if the minimum distance between a census tract in a MSA and a water body is greater than 5 miles we drop the distance to water body variable for workers located in that MSA. We find that for households with income less than $75,000, utility increases with distance from water, while for households with income above $75,000, utility decreases with distance from water. This means that, ceteris paribus, we should find higher income households locating near major water bodies.

There is evidence that the construction of the inter-state highway network in the U.S. was one factor behind the exodus of residents from U.S. central cities (Baum-Snow 2007). The inter-state network may have initially allowed workers to move to the suburbs by keeping their workplace in the central cities relatively accessible, but later allowed jobs to suburbanize too, thus in turn further increasing worker suburbanization as well. However, Table 4 reveals that workers across all household income categories prefer residence census tracts which are away from major highways. This apparent contradiction is easily explained by the fact that once we control for commuting time, being close to a major road probably has only negative effects. Such negative effects might arise from highway noise and the pollution and congestion associated with highway interchanges.

Table 4 also shows that workers prefer residence census tracts which are close to their place of work[8], but at the same time have less jobs in them. The reason for including distance in addition to travel time in the utility function is that distance might capture monetary costs of travel which are usually positively correlated with distance to work. For example, travelling an hour in Los-Angeles might involve travelling longer distances and higher monetary costs than, say, in New York. A large number of jobs in a residence census tract might create negative effects from commercial activity, and this would explain the negative sign.

We find that as the median household income of a residence census tract increases utility goes up for all workers, except the ones belonging to the poorest households. The negative effect of household income on the poorest households might be explained by the preference of the poorest to interact with others of similar incomes, or negative effects faced from explicit or implicit policies that attempt to keep poorer households from locating in affluent neighborhoods. An example of such a policy are large lot zoning laws which set a minimum lot size requirement

[8] Distance is measured as the straight line distance between the centroids of the work-residence census tract pair.

in suburban communities, thereby making it more cumbersome for poorer households to locate in such residential areas. An implicit effect might exist because of racial and ethnic prejudices by the more affluent groups. As many low income households also happen to belong to the minority community, the effect of such prejudices cannot be overlooked. Locating in affluent neighborhoods also imposes a higher tax burden on households, and hence may make poorer households avoid well off neighborhoods other things being equal.

As expected, workers across all household income groups prefer lower residential densities. We include the census tract's land area because they differ widely across, and even within MSAs'. Census tracts in the eastern part of the U.S., central cities of MSAs', are usually smaller, and are associated with higher population density. Therefore, the positive sign for the land area coefficient picks up both workers' desire for lower residential density.

Workers also prefer residence census tracts which have a larger number and newer residential houses, lower housing costs, and higher average number of bedrooms. We interacted the variables "age of house[9]" and "mean number of bedrooms" with the log of average earnings observed at the worker's place of work. This interaction allows these parameters to vary spatially and across income groups and by workplace, and also implies that the marginal utility of a worker for housing quality (measured either by age or bedrooms), increases with the worker's earnings.

Note that, the coefficient for total housing units is always positive and between zero and one. This coefficient can be interpreted as an inclusive value parameter from an additional level of nesting, where the worker chooses a particular housing unit within a census tract. Since, in

[9] "Age of house" = 2000 – "median year built" in Census tract.

our data, differences among housing units in the same tract are not observed, the expected utility from this final level of nesting is simply ln(*Total housing units*). Also, recall that housing costs per household was converted to housing costs per worker by dividing rent with the average number of workers per household. This conversion better reflects the housing cost of a worker, which is important since we are modeling the choice process at the level of a worker. Finally, the average number of bedrooms enters the utility function as a polynomial showing that utility increases with the number of bedrooms but at a decreasing rate. This result might arise from the fact that more bedrooms are associated with higher maintenance costs for larger houses.

The categorical variables *dumccr* and *dumccw* denote the location in the central city of the place of residence and place of work respectively. For workers' working in a central city, the variable $\exp(dumccr + dumccw) * \log(Earnings)$ is equal to $exp(2) * \log(Earnings)$ if they also reside in a central city, and it equals to $\exp(1) * \log(Earnings)$ if they reside outside a central city. The variable is equal to $\log(Earnings)$, if the worker both works and resides outside a central city (that is in suburban areas). The negative estimated sign for this variable implies that the suburban workers also most-prefer to live in suburban areas. Note, however, that although both central city and non-central city workers prefer to live outside a central city, the non-central city workers will enjoy a higher level of utility. So workers in general, all other things being equal, prefer to live outside a central city, regardless of where they work, but we find that getting a job outside a central city improves their utility even more. The interaction with log(*Earnings*) implies that the effect of workers' job location on their utility increases with their earnings, that is the higher earners value living and working in the suburbs more strongly than the lower earners.

We use the specification presented in Table 4 as a benchmark and estimate the nested logit model separately for each of the 275 MSAs present in our sample. We do this estimation from samples which pool workers across all household incomes for each MSA. It is impossible to discuss every MSA estimation result, so in Table 5 we present the results for seven of the largest MSAs spread across different geographical regions of the continental U.S.

The ranking of commuting modes vary across the cities in Table 5. In particular, we find that in Chicago, public transportation is the most preferred mode of commuting, while in New York and Boston public transportation comes second after driving to work. The popularity of public transportation in these cities is probably due to the presence of better public transportation infrastructure, and higher population density which makes for better accessibility to public transportation.

The inclusive value parameter lies between zero and one for all the cities, pointing to the fact that the nested logit model is consistent with utility maximization for all the cities. The inclusive value parameter is lowest for New York and highest for Miami, which indicates that the degree of correlation across commuting mode alternatives within the same work-residence census tract pair is highest in New York and lowest in Miami.

We again find a U-shaped relationship between utility and access to consumption opportunities in all the cities except Miami. In Miami this relationship is an inverted-U. The relationship between utility and distance to water also differs across cities. For New York, Los Angeles and Washington DC, utility increases with distance to water. For Chicago, Boston, Houston, and Miami utility decreases with distance to water.

Finally, the sign for the coefficient of $exp(dumccr)exp(dumccw)$*log(Earnings) is again negative for all the cities, except New York. This means that in New York, working and living in the central city (the five boroughs) is the most preferred option. All the other results are the same as the corresponding nation-wide estimation results, but one anomaly is that utility is decreasing in the number of rooms, in Los Angeles and Chicago.

The nested logit model estimates for the four income groups presented in Table 4 might suffer from bias since their samples were subject to censoring. However, comparing those parameter estimates with the estimates for all income groups whose sample was not subject to censoring does not reveal any significant bias.

6. Elasticity

The estimation results presented in Tables 4 and 5 can be used to compute choice elasticities. One use of these elasticities is to help calibrate discrete choice based models of land use such as the one discussed in Anas and Liu (2007). These elasticities might be of interest not only to economists, but to planners and policy makers as well. Of particular interest are the choice elasticity with respect to commuting time from the commuting mode choice model, and the elasticity with respect to housing cost from the residential location model.

Table 6 shows the U.S. MSA average own[10] and cross spot elasticities of commuting mode choice with respect to commuting times from the estimated nation-wide mode choice model. We find that on average for carpool, other modes and transit an approximately 1% increase in commuting times will reduce mode usage by approximately 0.16%. However, for drove alone, on average a 1% increase in commuting times will reduce usage by only 0.04%. If we increase the commuting time for carpool, other modes or transit by 1% then the effect on usage for the rest of the modes are modest, around 0.01%. However, increasing commuting time for drove alone by 1%, causes usage of carpooling, other modes and transit to increase by around 0.13%.

The highly inelastic demand for drove alone is probably because of the absence of other commuting mode choices from the residence locations of a large number of workers in the U.S., and also because driving alone allows for a more comfortable, private and safer commute than do the other three modes. The inelastic mode choice elasticities reported in Table 6 also results from the relatively high mode share (close to one) for drove alone, low shares (close to zero) for carpool, other modes, and transit in U.S. cities, and the sigmoid shape of the logit probability. The sigmoid shape implies that the logit probability responds little to changes in utility when the initial probability is close to zero or one, as compared to when the initial probability is around one-half (Train, 2009).

---

[10] Own price elasticity = $-0.18\frac{\sum_{MSA}(1-P_{m|MSA})N_{m|MSA}}{\sum_{MSA}N_{m|MSA}}$. $P_{m|MSA}$ is the mode share in an MSA and $N_{m|MSA}$ is the number of workers using mode m in that MSA. Cross price elasticity = $-0.18\frac{\sum_{MSA}(P_{m|MSA})N_{m|MSA}}{\sum_{MSA}N_{m|MSA}}$.

Comparing the mode choice elasticities reported in this paper with the existing literature is complicated by the fact that studies differ temporally, spatially, as well as in data and model specification. In addition, given the huge literature on commuting mode choice it is difficult to compare our results with all existing studies. However, Chan and Ou (1978) and Litman (2013) provide nice summaries of travel demand elasticities. Chan and Ou (1978) report an average own time elasticity for auto and bus of about -0.6, which is significantly higher than the ones reported in Table 6. This difference might arise because most of the elasticities tabulated by Chan and Ou (1978) are old, and there is evidence that U.S. transport demand elasticities have declined significantly during the last half of the twentieth century (Litman, 2013). A major reason for the decline in elasticity could be the fact that jobs and residences have continued to decentralize into the suburbs where public transit is less available and less accessible.

Table 6 also presents own and cross commuting mode choice elasticity estimates from some recent studies surveyed in Litman (2013). TRACE (1999) was a research program carried out by a group of European consultants and Universities with the aim of producing a handbook of both short and long term car travel demand elasticities with respect to cost and time. The TRACE program covered mostly European countries. We see that according to TRACE, a 1% increase in car commuting times in Europe will reduce car ridership by 0.41%, which is approximately ten times higher than our estimate for the U.S., but closer to the estimate reported by Chan and Ou (1978). A 1% increase in car commuting times will increase transit ridership by 2%, and the slow mode usage by 0.25%. These estimates are around 16 and 2 times higher than what we find in this paper for the U.S. In comparing the TRACE estimates with our estimates we must keep in mind that European cities are more compact (allowing for more frequent usage of slow modes like riding a bike or walking to work) and have better public transportation infrastructure than most U.S. cities. This implies that European workers have more options when choosing their commuting mode than their U.S. counterparts. In other words, the greater availability of substitutes for car as a mode of commuting yields higher elasticities as compared to our study.

According to elasticity estimates from Dowling Associates (2005) presented in Litman (2013) a 1% increase in mode specific travel times during the morning peak hours in Portland Oregon should lower car usage by 0.22%, shared ride usage by 0.30% and transit usage by 0.12%. Except for the car choice elasticity, the own elasticity estimates reported by Dowling Associates (2005) are reasonably close to our estimates. We find that a 1% increase in car travel time should reduce carpool usage by 0.14% which is significantly higher than the 0.03% found by Dowling Associates (2005). The reason for this difference might be that in our case we consider the effect of changing carpool commute times; holding drove alone commute time fixed. While in reality changing carpool times will also change drove alone times and hence the elasticity would be lower. All our other cross elasticities, however, closely match those reported by Dowling Associates (2005).

Frank et.al (2007) study tour[11] mode choice for the Seattle region using data from the 1999 Puget Sound Regional Council's Household Travel Survey. They find that for home based work tours increasing auto travel times by 1% should reduce drove alone usage by 0.03%, which is remarkably close to our own estimate of 0.04%. They also find that increasing auto travel time by 1% will increase transit usage by 0.31%, bike usage by 0.28% and walking by 0.05%. Again,

[11] A tour in transportation literature is a sequence of individual trips (trip chain) that usually starts and ends at home.

these cross elasticities are close to our own findings given the fact that we lump together bike and walking into one single mode. Their own and cross elasticities with respect to transit travel times also match our findings. Given that Frank et.al (2007) uses micro data from 1999, same period as in this study, and find elasticity estimates reasonably close to our own estimates implies that the use of aggregated data does not result in large biases. This is similar to the conclusion reached by Anas (1981).

The result from Table 4 implies a residential location choice elasticity[12] with respect to housing cost of about -0.28. Table 4 also indicates that there is an inverted U shaped relationship between housing cost elasticity and workers household income category. This might be due to the fact that workers who belong to the lowest and highest household income categories are more influenced by unobserved idiosyncratic factors in making their choices, than are workers who belong to the middle household income categories.

The only studies directly comparable to ours were done by Anas (1981) and Anas and Chu (1984). The latter reports a range for residential location choice elasticity of -0.26 to -0.86 depending on their model specification, and finds that this range included residential location choice elasticity estimates from other studies. Our residential location choice estimates from the different national samples all fall within this range. In particular, for Chicago, we find a housing cost elasticity of -0.19 as compared to an average of -0.36 by Anas and Chu (1984). This

[12] $\beta_k(1 - P_{j|i,s})$ is the housing cost elasticity for workers, where β_k is the coefficient for log(*Annual adj rent*). Since the number of census tracts is large, we can fairly safely assume $1 - P_{j|i,s} \cong 1$. In that case the housing cost elasticity would be approximately β_k.

difference might be accounted for by the fact that we consider commutes to all sections of the MSA, whereas they included only commuters to the downtown.

The mode choice elasticities we get from the national pooled sample (Table 4, Income all) vary across MSAs due to differences in MSA mode shares. However, the parameter in the mode choice elasticity has the same value of -0.18 across all MSAs. The residential location choice elasticity estimates from the same sample does not vary across MSAs. In order to get an idea of the distribution of these elasticities across MSAs we use the elasticities from the MSA specific estimations since in these estimations the parameters are allowed to vary across MSAs[13]. In Figure 3c we plot the distribution of the weighted average of the four own elasticities of mode choice with respect to commuting time, the weights being the mode shares in the MSA. In Figure 3d we plot the distribution of the residential location choice elasticity. We find that among a family of distributions, the Burr (Singh-Maddala) best explains the data generation process for the MSA elasticities. For comparison, in both figures we also plot the Kernel probability density estimates. As can be seen from the figures, the Burr and Kernel probability densities closely hug each other.

The average MSA own mode choice elasticity and residential location choice elasticity from the MSA specific estimations are, -0.07 and -0.22 respectively. These values are close to the corresponding elasticity estimates of -0.06 and -0.18 that we get from the national pooled

[13] The distributions of mode choice and residential location elasticities in Figure 3 are based on a sample of 211 MSAs for which we obtained the correct signs for both log(Mean commuting time) and log(Housing costs).

sample. Therefore, our average MSA elasticity estimates reported in Table 6 are not biased significantly by pooling workers across all MSAs.

The mode choice and residential location elasticities are closely related to workers' preferences[14]. Population density in a MSA is the outcome of many workers residential location choices based on these preferences. A natural question to ask then is: how are MSA population densities affected by these elasticities? In Table 7, models (1) and (2) show that increasing residential location choice elasticity lowers population density while the opposite result holds for mode choice elasticity. Both results are intuitive. If people become more sensitive to housing cost they are likely to move to the suburbs where housing costs is usually lower. If people are sensitive to commuting time as represented by a higher mode choice elasticity they are likely to move closer to their work location. The first effect will tend to lower density, the second increase density. We find that higher income lowers density while higher population increases density. Note that model 2 explains almost 50 percent of the variation in population density. The effect of income and population on density also agrees with the comparative static results from a standard monocentric theory model (Wheaton 1974).

Urban sprawl can be thought of as opposite to population density. In fact the correlation coefficient between our population density measure and the sprawl index reported in Burchfield et al. (2006) is -0.40. To check the robustness of our previous results, we regress the sprawl index on these elasticities. As expected, models (3) and (4) in Table 7 have the opposite signs to what we find in models (1) and (2). Higher residential location choice elasticity tends to increase

---

[14] These elasticities are also affected by other factors. For example, mode choice elasticity is affected by the availability of public transportation.

sprawl, higher mode choice elasticity tends to reduce sprawl. Higher income raises sprawl, higher population lowers sprawl. Since availability of public transportation raises mode choice elasticity more transit should raise density or reduce sprawl in MSA.

7. Conclusion

As we saw from the estimation results for the individual metropolitan areas, there exists a distribution across metropolitan areas for the parameters in the utility function, relative to the estimates from the nationwide sample. But given the large number of metropolitan areas, this chapter provides only a limited insight about the parameter distribution.

It will be useful if a more complete picture of the parameter distribution can be provided. This would require applying some different discrete choice model, like the mixed logit, at a higher level of aggregation. Mixed logit allows for random parameters. The estimates from the nationwide sample could be used as a starting point to estimate the parameter distribution across the MSAs.

Finally, an extension to this chapter can involve adding another layer to the nested logit model, where workers choose their residence metropolitan area. This model can be used to study factors that determine the inter-metropolitan migration of workers. It can also be used to create a metropolitan utility index. One can use such an index to examine the attractiveness of a metropolitan area, and how different policies affect the index.

8. References

[1] A. Anas and C. Chu, "Discrete choice models and the housing price and travel to work elasticities of location demand," Journal of Urban Economics, vol. 15, pp. 107-123, 1984.

[2] A. Anas, "A unified theory of consumption, travel and trip chaining," Journal of Urban Economics, vol. 62, pp. 162-186, 2007.

[3] A. Anas and Y. Liu, "A regional economy, land use, and transportation model (RELU-TRAN(c)): formulation, algorithm design, and testing," Journal of Regional Science, vol. 47, no. 3, pp. 415-455, 2007.

[4] Dowling Associates, "Predicting air quality effects of traffic flow improvements: final report and user guide, NCHRP report 535," Transportation Research Board, pp. 1-241, 2005.

[5] http://nationalatlas.gov, "National atlas of the United States".

[6] TRACE, "Final report for publication," 1999.

[7] "Census transportation planning package," http://www.transtats.bts.gov/.

[8] A. Anas, R. Arnott and K. A. Small, "Urban spatial structure," Journal of Economic Literature, vol. 36, pp. 1426-1464, 1998.

[9] A. Anas, "The estimation of multinomial logit models of joint location and travel mode choice from aggregated data," Journal of Regional Science, vol. 21, no. 2, pp. 223-242, 1981.

[10] A. Bento, M. L. Cropper, A. M. Mobarak and K. Vinha, "The effects of urban spatial structure on travel demand in the United States," The Review of Economics and Statistics, vol. 87, no. 3, pp. 466-478, 2005.

[11] N. Baum-Snow, "Did highways cause suburbanization?," The Quarterly Journal of Economics, vol. 122, no. 2, pp. 775-805, 2007.

[12] Y. Chan and F. Ou, "Tabulating demand elasticities for urban travel forecasting," Transportation Research Record, vol. 673, pp. 40-46, 1978.

[13] E. Glaeser, M. E. Kahn and J. Rappaport, "Why do the poor live in cities? The role of public transportation," Journal of Urban Economics, vol. 63, no. 1, pp. 1-24, 2008.

[14] E. Glaeser, J. Kolko and A. Saiz, "Consumer city," Journal of Economic Geography, vol. 1, pp. 27-50, 2001.

[15] T. Litman, "Understanding transport demand elasticities, how prices and other factors affect travel behavior," Victoria Transport Policy Institute, 2013.

[16] S. F. Leroy and J. Sonstelie, "Paradise lost and regained: transportation, innovation, and residential location," Journal of Urban Economics, vol. 13, pp. 67-89, 1983.

[17] F. Lawrence, M. Bradley, S. Kavage, J. Chapman and T. K. Lawton, "Urban form, travel time, and cost relationships with tour complexity and mode choice," Transportation, vol. 35, pp. 37-54, 2008.

[18] Y. M. Ioannides and J. E. Zabel, "Interactions, neighborhood selection and housing demand," Journal of Urban Economics, vol. 63, pp. 229-252, 2008.

[19] P. Mieszkowski and E. S. Mills, "The causes of metropolitan suburbanization," Journal of Economic Perspective, vol. 7, no. 3, pp. 135-147, 1993.

[20] J. M. Quigley, "Consumer choice of dwellings, neighborhood and public services," Regional Science and Urban Economics, vol. 15, pp. 41-63, 1985.

[21] S. Nerella and C. Bhat, "Numerical analysis of the effects of sampling of alternatives in discrete choice models," Transportation Research Record: Journal of the Transportation Research Board, No. 1984, Transportation Research Board of the National Academies, Washington D.C., pp. 11-19, 2004.

[22] C. Rapaport, "Housing demand and community choice: an empirical analysis," Journal of Urban Economics, vol. 42, pp. 243-260, 1997.

[23] K. Small and E. Verhoef, The economics of urban transportation, Routledge, 2007.

[24] K. Train, Discrete choice methods with simulations, 2 ed., Cambridge: Cambridge University Press, 2009.

[25] P. M. Wadell, C. Bhat, N. Eluru, L. Wang and R. M. Pendyala, "Modeling the interdependence in household residence and workplace choices," Transportation Research Record: Journal of the Transportation Research Board, 2003, pp. 84-92, 2007.

[26] J. Yates and D. F. Mackay, "Discrete choice modelling of urban housing markets: a critical review and an application," Urban Studies, vol. 43, no. 3, pp. 559-581, 2006.

[27] J. Woolridge, Econometric analysis of cross section and panel data, MIT press, 2010.

[28] M. White, "Location choice and commuting behavior in cities with decentralized employment," Journal of Urban Economics, vol. 24, no. 2, pp. 129-152, 1988.

[29] M. Burchfield, H. G. Overman, D. Puga and M. A. Turner, "Causes of sprawl: A potrait from space," The Quarterly Journal of Economics, vol. 121, no. 2, pp. 587-633, 2006.

[30] W. C. Wheaton, "A comparative static analysis of urban spatial structure," Journal of Economic Theory, vol. 9, no. 2, pp. 223-237, 1974.

9. Figures and Tables

Table 1: Shares of commuting patterns by mode, year 2000

Workplace	Residence	All modes	Drove	Carpool	Transit	Other modes
Suburb	Suburb	43.15	45.62	39.89	9.23	49.86
Suburb	Central city	9.12	9.37	11.62	7.17	3.45
Central city	Suburb	19.84	21.05	19.49	23.00	3.38
Central city	Central city	27.89	23.97	29.00	60.60	43.31

Source: CTPP 2000 (Note: columns sum to 100%)

Table 2: Data sources

Variable	Source
Accessibility to consumption	CTPP2 2000, CTPP3 2000
Distance to work	U.S Census 2000 SF3
Distance to road	National Atlas of the U.S (road layer)
Distance to water	National Atlas of the U.S (hydrography layer)
Flow of workers between work and home	CTPP3 2000
Land	U.S Census 2000 SF3
Mean number of workers per household	CTPP1 2000
Mean commute time	CTPP3 2000
Median year of construction	U.S Census 2000 SF3
Median household income	U.S Census 2000 SF3
Median number of bed rooms	U.S Census 2000 SF3
Housing cost	U.S Census 2000 SF3
Residence/Work central city dummy	U.S Census Bureau central city boundary map layer
Retail, arts, entertainment jobs	CTPP2 2000
Total number of housing units	U.S Census 2000 SF3
Total number of jobs	CTPP2 2000
Earnings	CTPP2 2000

Table 3a: Descriptive statistics of Census tracts

	Census Division					
Variables	New England		Middle Atlantic		East North Central	
	Mean	Std	Mean	Std	Mean	Std
Accessibility to consumption*	0.25	0.67	0.5	1.78	0.27	0.72
Distance to road***	2.12	10.64	2.46	14.39	3.03	16.91
Distance to water*	15.79	78.75	12.96	59.65	29.34	133.31
Distance to work***	8.38	42.06	8.4	37.36	8.3	36.11
Earnings**	35360	28660	34728	28251	32963	29306
Earnings, income group 1**	14353	6882	14151	7002	14243	7721
Earnings, income group 2**	25371	13374	24958	14089	24983	14638
Earnings, income group 3**	33427	20720	33324	22155	33135	25122
Earnings, income group 4**	46382	34411	46650	35137	45066	36585
Land area*	6.23	43.12	6.42	73	10.54	112.2
Mean number of bedrooms*	2.67	1.95	2.64	2.23	2.72	1.82
Mean travel time, carpool***	25.46	73.35	29.67	79.55	26.09	70.91
Mean travel time, drove alone***	23.35	74.12	24.91	74.70	22.95	70.64
Mean travel time, other modes***	13.73	97.16	14.73	84.52	13.53	97.50
Mean travel time, transit***	46.68	102.68	51.81	103.02	48.01	104.20
Mean workers per household*	1.66	0.57	1.63	0.61	1.63	0.62
Mean workers per household*, income group 1	1.2	0.42	1.21	0.42	1.21	0.46
Mean workers per household*, income group 2	1.42	0.58	1.44	0.62	1.45	0.6
Mean workers per household*, income group 3	1.72	0.71	1.72	0.82	1.75	0.74
Mean workers per household*, income group 4	1.98	1.09	1.95	1.15	1.98	0.98
Median year of construction*	1958	65	1957	57	1964	68
Median household income*	56587	106456	52832	95811	50510	86530
Annual housing cost*	15731	44456	13630	35613	11475	27599
Total housing units*	2142	3578	2063	4471	2081	4376
Total jobs*	2414	15179	2022	13131	2216	16051

*Attribute varies by residence location
**Attribute varies by work location
***Attribute varies by residence and work location

Table 3b: Descriptive statistics of Census tracts

	Census Division					
Variables	West North Central		South Atlantic		East South Central	
	Mean	Std	Mean	Std	Mean	Std
Accessibility to consumption*	0.36	0.78	0.24	0.64	0.29	0.61
Distance to road***	3.44	31.06	4.03	28.77	4.29	34.17
Distance to water*	18.54	150.01	46.52	270.28	40.27	216.53
Distance to work***	7.56	36.61	8.67	40.83	7.97	39.08
Earnings**	31614	28984	32011	32914	29667	28072
Earnings, income group 1**	14418	8169	14417	7917	14204	7892
Earnings, income group 2**	24314	14866	24219	14722	24137	14040
Earnings, income group 3**	32130	22633	32403	22335	32122	21763
Earnings, income group 4**	44358	32377	46059	34317	44722	30252
Land area*	19.71	359.79	12.93	191.98	20.76	208.96
Mean number of bedrooms*	2.73	2.1	2.72	2.54	2.74	2.15
Mean travel time, carpool***	23.40	62.19	28.73	81.50	25.54	73.81
Mean travel time, drove alone***	20.53	61.30	24.98	78.68	22.07	70.72
Mean travel time, other modes***	13.10	111.56	16.72	128.12	15.57	132.46
Mean travel time, transit***	37.99	90.98	48.18	108.97	41.91	109.42
Mean workers per household*	1.66	0.66	1.62	0.68	1.58	0.66
Mean workers per household*, income group 1	1.23	0.45	1.22	0.43	1.21	0.36
Mean workers per household*, income group 2	1.51	0.64	1.49	0.64	1.5	0.56
Mean workers per household*, income group 3	1.82	0.72	1.77	0.83	1.78	0.74
Mean workers per household*, income group 4	2.01	0.95	1.93	1.09	1.93	1.02
Median year of construction*	1968	78	1976	66	1974	73
Median household income*	50103	89125	49585	106443	42942	94327
Annual housing cost*	10351	23461	11210	31405	9074	24383
Total housing units*	2047	4161	2614	6598	2360	6138
Total jobs*	2282	16611	2583	19491	2184	16513

*Attribute varies by residence location
**Attribute varies by work location
***Attribute varies by residence and work location

Table 3c: Descriptive statistics of Census tracts

	Census division					
Variables	West South Central		Mountain		Pacific	
	Mean	Std	Mean	Std	Mean	Std
Accessibility to consumption*	0.3	1	0.37	0.78	0.27	0.86
Distance to road***	4.42	37.79	3.2	25.7	3.87	96.54
Distance to water*	27.56	118.87	33.74	155.59	19.36	110.45
Distance to work***	8.74	40.97	7.73	41.83	9.53	47.26
Earnings**	30526	31528	30953	28451	33637	29685
Earnings, income group 1**	13787	8653	14067	9144	13432	7698
Earnings, income group 2**	23953	14832	24048	14150	24400	14397
Earnings, income group 3**	32630	22689	32221	20958	33494	22148
Earnings, income group 4**	46498	34689	44568	33241	47050	34098
Land area*	17.21	310.32	32.65	1150.44	14.34	700.44
Mean number of bedrooms*	2.62	2.39	2.75	2.76	2.55	2.41
Mean travel time, carpool***	26.96	76.25	25.68	69.37	29.44	79.33
Mean travel time, drove alone***	23.41	73.00	21.95	66.08	25.10	76.05
Mean travel time, other modes***	17.06	136.55	16.03	124.30	16.23	115.97
Mean travel time, transit***	47.64	115.46	45.18	104.84	47.53	105.95
Mean workers per household*	1.59	0.64	1.64	0.73	1.65	0.72
Mean workers per household*, income group 1	1.25	0.45	1.26	0.58	1.27	0.54
Mean workers per household*, income group 2	1.53	0.77	1.53	0.75	1.5	0.82
Mean workers per household*, income group 3	1.78	0.92	1.8	0.87	1.73	1.01
Mean workers per household*, income group 4	1.91	1.13	1.99	1.25	1.94	1.36
Median year of construction*	1975	62	1977	67	1969	59
Median household income*	45377	100735	48902	93880	53606	99869
Annual housing cost*	8432	23593	12537	31431	18071	53461
Total housing units*	2299.34	5277	2127	4823	2093	4114
Total jobs*	2212	19058	2151	16926	2230	18089

*Attribute varies by residence location
**Attribute varies by work location
***Attribute varies by residence and work location

Table 4: Estimation results by income group (All MSAs pooled)

Independent variables	Income all	Income group 1	Income group 2	Income group 3	Income group 4
Carpool dummy	-1.25 (-3218.29)	-0.73 (-641.64)	-0.84 (-831.38)	-0.93 (-951.90)	-1.11 (-1334.66)
Public transportation dummy	-0.90 (-1230.55)	-0.52 (-259.06)	-0.60 (-274.53)	-0.62 (-271.25)	-0.52 (-338.80)
Other modes dummy	-0.78 (-1477.43)	-0.32 (-252.20)	-0.55 (-403.10)	-0.58 (-414.80)	-0.40 (-365.57)
log(Mean commute time)	-0.18 (-515.60)	-0.10 (-109.17)	-0.13 (-133.48)	-0.14 (-148.20)	-0.23 (-306.16)
Inclusive value	0.63 (1749.81)	0.89 (845.12)	0.88 (856.30)	0.86 (865.66)	0.77 (930.42)
Accessibility to consumption	-0.94 (-427.54)	-0.89 (-80.98)	-0.93 (-119.45)	-1.06 (-99.87)	-1.08 (-233.11)
Accessibility to consumption sq	0.25 (255.01)	0.16 (32.90)	0.14 (39.47)	0.21 (49.09)	0.26 (135.34)
Distance to water	-0.0004 (-11.84)	0.002 (15.94)	0.001 (14.15)	0.000004 (0.04)	-0.001 (-19.08)
Distance to water sq	0.00004 (64.08)	-0.00001 (-4.87)	0.000008 (3.68)	0.00002 (13.98)	0.0001 (42.73)
log(Distance to road)	0.05 (236.94)	0.03 (52.88)	0.02 (43.73)	0.02 (54.53)	0.04 (92.03)
log(Distance to work)	-0.76 (-4817.38)	-0.43 (-842.54)	-0.45 (-991.03)	-0.47 (-1095.78)	-0.54 (-1398.42)
log(Total jobs)	-0.05 (-467.54)	-0.01 (-48.77)	-0.04 (-124.86)	-0.04 (-154.76)	-0.04 (-159.47)
log(Median household income)	0.35 (467.15)	-0.23 (-106.20)	0.40 (174.13)	0.77 (335.36)	0.95 (384.93)
log(Land area)	0.08 (709.67)	0.01 (35.33)	0.03 (106.14)	0.04 (148.51)	0.04 (186.42)
log(Total housing units)	0.68 (2188.39)	0.59 (567.45)	0.62 (681.20)	0.64 (715.61)	0.68 (1006.96)
log(Annual adjusted housing cost)	-0.28 (-514.79)	-0.20 (-113.42)	-0.44 (-279.86)	-0.52 (-369.43)	-0.13 (-102.43)
Median age of house*log(Earnings)	-0.0005 (-453.85)	-0.0004 (-105.60)	-0.0004 (-122.09)	-0.0005 (-183.99)	-0.0006 (-246.14)

	Continued				
Mean number of bedrooms*log(Earnings)	0.01 (92.49)	0.01 (22.02)	0.04 (75.92)	0.06 (119.95)	0.04 (124.78)
Mean number of bedrooms sq*log(Earnings)	-0.002 (-59.44)	-0.004 (-30.51)	-0.01 (-94.30)	-0.01 (-123.14)	-0.006 (-97.90)
exp(DUMCCR+DUMCCW)*log(Earnings)	-0.001 (-169.36)	-0.0002 (-8.17)	-0.0008 (-38.51)	-0.001 (-53.80)	-0.0006 (-37.74)
$\boldsymbol{\rho_m^2}$	0.53	0.29	0.39	0.44	0.44
$\boldsymbol{\rho_j^2}$	0.61	0.55	0.48	0.46	0.55
Census Tract Pairs	4,372,582	620,673	770,735	850,328	926,848
Total Workers	95,854,466	10,912,161	14,080,682	16,355,320	23,581,160

T-statistics are in parenthesis
Income Group 1: Less than $30,000
Income Group 2: $30,000-$4,999
Income Group 3: $50,000-$74,999
Income Group 4: $75,000 and above
$\rho^2 = \frac{ll'-ll(0)}{ll^*-ll(0)}$. ll' is the log likelihood function evaluated at the estimated probabilities, ll^* is the log likelihood function evaluated at the observed probabilities (the best possible value of the log likelihood), and ll^0 is the log likelihood function when parameters are set to zero (the log likelihood value of the null hypothesis).

Table 5: Estimation results by selected MSAs (All income groups pooled)

Independent Variables	New York	Los Angeles	Chicago	Washington, DC	Boston	Houston	Miami
Carpool dummy	-0.95 (-554.63)	-0.79 (-492.59)	-1.03 (-465.35)	-1.04 (-497.23)	-1.25 (-463.50)	-1.08 (-416.12)	-0.97 (-319.20)
Public transportation dummy	-0.26 (-129.65)	-0.69 (-208.72)	0.03 (12.23)	-0.51 (-174.77)	-0.15 (-45.33)	-1.05 (-199.43)	-1.00 (-170.35)
Other modes dummy	-0.26 (-141.38)	-0.25 (-114.42)	-0.40 (-140.55)	-0.41 (-142.13)	-0.52 (-171.18)	-0.71 (-168.78)	-0.49 (-105.41)
log(Mean commute time)	-0.26 (-206.66)	-0.24 (-169.41)	-0.22 (-120.01)	-0.22 (-117.71)	-0.24 (-113.47)	-0.12 (-46.66)	-0.17 (-58.23)
Inclusive value	0.73 (752.89)	0.75 (613.73)	0.77 (476.47)	0.78 (448.20)	0.72 (333.67)	0.76 (273.88)	0.83 (284.44)
Accessibility to consumption	-0.44 (-146.25)	-5.20 (-129.07)	-1.86 (-95.62)	-3.55 (-92.71)	-2.11 (-68.58)	-4.37 (-67.53)	0.76 (10.47)
Accessibility to consumption sq	0.06 (45.14)	8.74 (58.78)	1.69 (53.19)	2.86 (34.95)	0.76 (13.66)	5.97 (35.93)	-1.79 (-13.34)
Distance to water	0.001 (9.59)	0.0004 (3.91)	-0.005 (-26.98)	0.0007 (3.88)	-0.005 (-22.02)	-0.004 (-11.74)	-0.0001 (-0.16)
Distance to water sq	0.0001 (18.69)	0.00001 (10.01)	0.0002 (47.37)	-0.00002 (-6.80)	0.0001 (13.00)	0.0001 (9.64)	0.0003 (4.83)
log(Distance to road)	0.04 (61.98)	0.02 (27.22)	0.01 (12.93)	0.05 (52.10)	0.05 (38.40)	-0.003 (-2.69)	0.09 (55.67)
log(Distance to work)	-0.56 (-1019.21)	-0.53 (-881.42)	-0.64 (-793.22)	-0.70 (-791.79)	-0.81 (-808.81)	-0.73 (-660.75)	-0.64 (-457.72)
log(Total jobs)	-0.02 (-62.77)	-0.02 (-55.75)	-0.04 (-66.47)	-0.03 (-60.78)	-0.03 (-47.27)	-0.03 (-36.93)	-0.01 (-11.80)
log(Median household income)	0.24 (129.87)	0.27 (90.16)	0.28 (81.20)	0.33 (72.73)	0.38 (75.27)	0.40 (70.87)	0.28 (46.32)
log(Land area)	0.03 (63.99)	0.03 (57.77)	0.04 (53.40)	0.08 (117.39)	0.10 (93.51)	0.02 (23.33)	0.03 (27.84)
log(Total housing units)	0.53 (491.45)	0.57 (421.42)	0.60 (366.48)	0.71 (395.79)	0.74 (338.34)	0.75 (311.76)	0.51 (174.83)
log(Annual adjusted housing cost)	-0.13 (-89.47)	-0.18 (-96.01)	-0.19 (-70.80)	-0.24 (-70.69)	-0.28 (-69.10)	-0.30 (-78.57)	-0.18 (-39.12)
Median age of house*log(Earnings)	-0.0002 (-43.33)	-0.0004 (-81.75)	-0.0002 (-29.48)	-0.0005 (-87.72)	-0.0001 (-16.01)	-0.0009 (-95.55)	-0.0005 (-41.30)
Mean number of bedrooms*log(Earnings)	0.008 (17.35)	-0.007 (-11.90)	-0.003 (-2.89)	0.008 (11.43)	0.001 (0.97)	0.006 (6.59)	0.03 (22.65)
Mean number of bedrooms sq*log(Earnings)	-0.001 (-11.60)	0.0019 (18.3221)	0.001 (7.14)	-0.001 (-9.29)	0.0003 (1.20)	0.0001 (0.61)	-0.003 (-14.12)
exp(DUMCCR+DUMCCW)*log(Earnings)	0.001 (55.76)	-0.0006 (-21.51)	-0001 (-1.38)	-0.002 (-41.62)	-0.001 (-27.80)	-0.0009 (-16.98)	-0.001 (-19.85)
$\boldsymbol{\rho_m^2}$	0.41	0.38	0.37	0.40	0.44	0.52	0.49
$\boldsymbol{\rho_j^2}$	0.53	0.54	0.58	0.63	0.68	0.60	0.56

Continued

Census Tract Pairs	594,758	421,334	238,833	183,904	135,746	99,812	71,584
Total Workers	8,948,881	6,416,880	4,067,628	3,536,658	2,738,903	1,974,126	1,532,630

T-statistics are in parenthesis

Table 6: Own and cross elasticity of commuting mode choice with respect to commuting times

	Carpool	Drove alone	Other modes	Transit
CARPOOL	−0.16♦	0.02*	0.02*	0.02*
	−0.30[2]	0.03[2]		0.03[2]
DROVE ALONE	0.14*	−0.04*	0.13*	0.12*
		−0.41[1]	0.25[1]	2.00[1]
	0.03[2]	−0.22[2]		0.01[2]
		−0.03[3]	0.28[3],0.05[3]	0.31[3]
OTHER	0.01*	0.01*	−0.17*	0.01*
TRANSIT	0.01*	0.01*	0.01*	−0.16*
	0.03[2]	0.03[2]		−0.12[2]
	0.03[3]	0.02[3]	0.08[3],0.01[3]	−0.39[3]

♦A 1% increase in carpool commuting time will reduce carpool usage by 0.16%, increase drove alone, other modes and transit usage by 0.02%.
*=Our estimate,1=TRACE (1999) , 2=Dowling Associates (2005), 3=Frank et.al (2008)
*Source: Understanding Transport Demands Elasticities, How Prices and Other Factors Affect Travel Behavior, Victoria Transport Policy Institute, 2013

Table 7: Effect of mode choice elasticity and housing cost elasticity on urban population density

	Population density		Sprawl index	
	(1)	(2)	(3)	(4)
Housing cost elasticity	-0.51 (-3.86***)	-0.19 (-1.62)	0.17 (3.24**)	0.09 (1.75.)
Mode choice elasticity	7.31 (4.09***)	5.30 (4.65***)	-1.76 (-3.56***)	-1.46 (-3.14**)
log(Average income)		-0.22 (-0.81)		0.23 (1.64)
log(Total population)		0.28 (9.81***)		-0.07 (-4.74***)
$\mathbf{R^2}$	0.20	0.47	0.08	0.16
Number of observations	211	211	211	211

Unit of observation is MSA, 2000 geography.
Heteroscedastic robust standard errors in parenthesis.
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 1: Major roads and water bodies

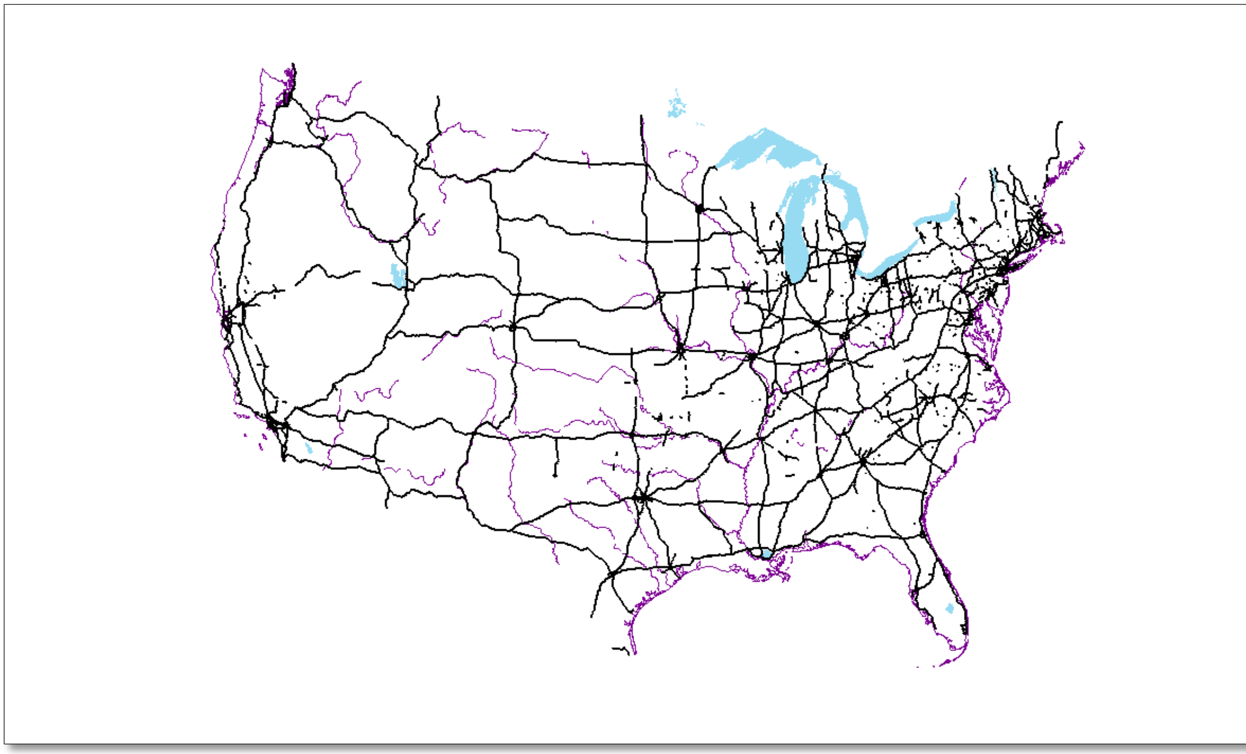


Figure 2 : Nesting structure

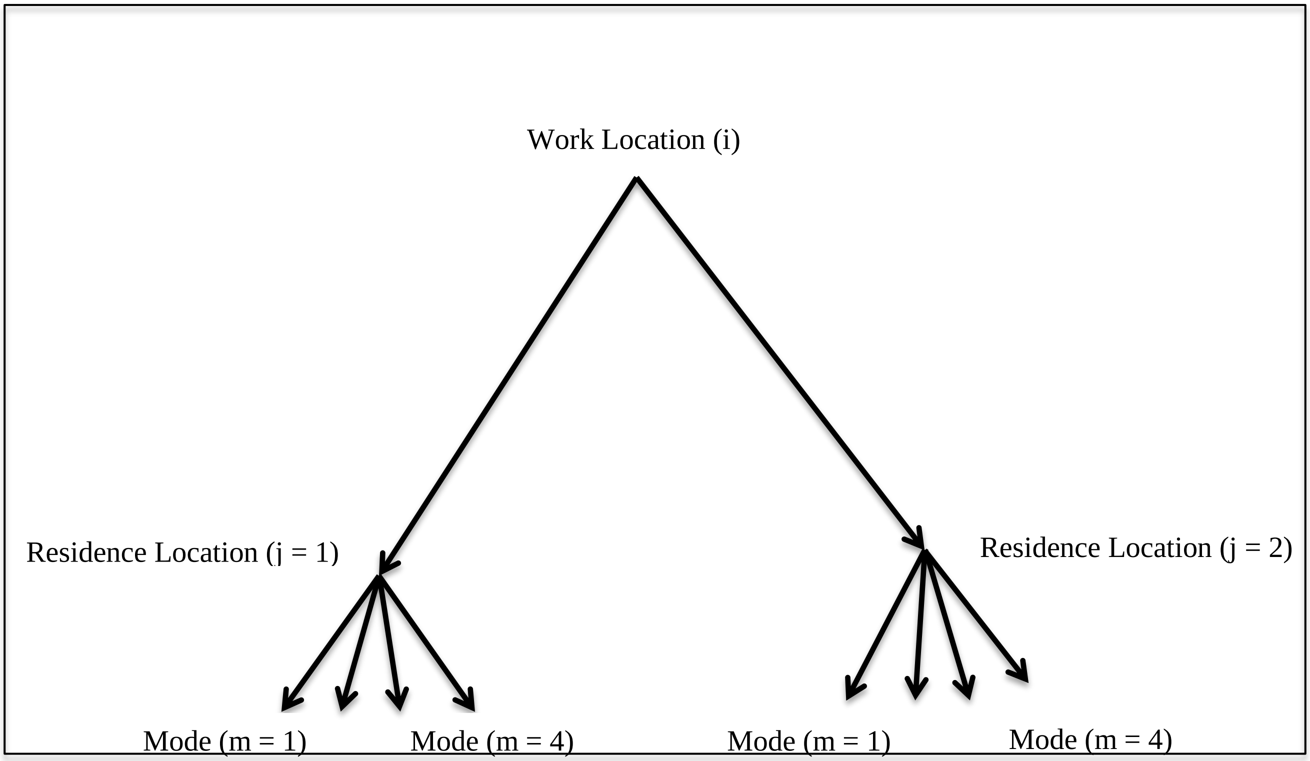
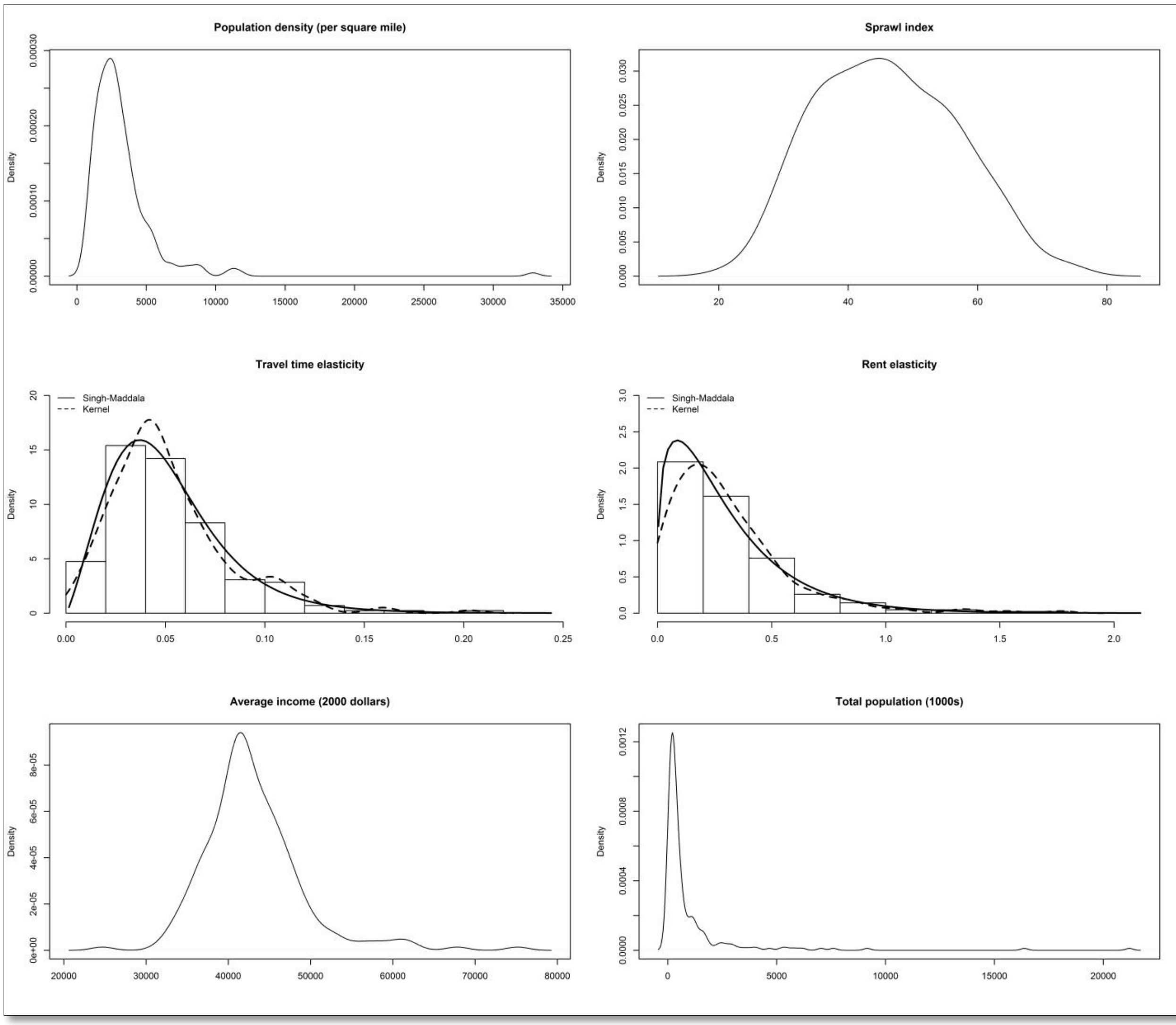


Figure 3: Data distributions



$\text{Population density}_{\text{MSA}} = \frac{\sum_{\text{i,MSA}} \text{population}_{\text{i,MSA}} \left(\frac{\text{population}_{\text{i,MSA}}}{\text{land}_{\text{i,MSA}}} \right)}{\sum_{\text{i}} \text{population}_{\text{i,MSA}}}$, i = census tract

Sprawl index = Sprawl measure 1992*

Mode choice elasticity = $\sum_m \delta_{mtt}(1 - P_m)\, P_m$, δ_{mtt} = Coefficient of log mean travel time, P_m = observed mode share

$\text{Average income}_{\text{MSA}} = \frac{\text{Aggregate household income}_{\text{MSA}}}{\text{Total number of workers}_{\text{MSA}}}$

*Source: “Causes of sprawl: A portrait from space, HG Overman, M Burchfield, D Puga, M Turner, Quarterly journal of economics 121 (2), 587-533, 2006”

The pdf of the Burr (Singh-Maddala) distribution is $\text{f(x)} = \frac{\alpha\gamma\left(\frac{x}{\theta}\right)^{\gamma}}{x\left(1+\left(\frac{x}{\theta}\right)^{\gamma}\right)^{\alpha+1}}$

The parameters of the Burr (Singh-Maddala) distribution for the:
mode choice elasticity: $\alpha = 2.98, \gamma = 2.22, \theta = 0.08$
housing cost elasticity: $\alpha = 8.16, \gamma = 1.28, \theta = 1.56$