A JOINT MODEL OF RESIDENTIAL LOCATION AND BICYCLE OWNERSHIP: ACCOUNTING FOR SELF-SELECTION AND UNOBSERVED HETEROGENEITY

Abdul Rawoof Pinjari
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: (512) 964-3228; Fax: (512) 475-8744; Email: abdul.pinjari@mail.utexas.edu

Naveen Eluru
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: (512) 471-4535; Fax: (512) 475-8744; Email: naveeneluru@mail.utexas.edu

Chandra R. Bhat*
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: (512) 471-4535; Fax: (512) 475-8744; Email: bhat@mail.utexas.edu

Ram M. Pendyala
Arizona State University
Department of Civil and Environmental Engineering
Room ECG252, Tempe, AZ 85287-5306
Tel: (480) 727-9164; Fax: (480) 965-0557; Email: ram.pendyala@asu.edu

Erika Spissu
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: (512) 232-6599; Fax: (512) 475-8744; Email: espissu@unica.it

*corresponding author

ABSTRACT

This paper presents a joint model of residential location choice and bicycle ownership with the intent of disentangling the true causal effects of the built environment on household bicycle ownership from spurious associative effects. The issue at stake is whether the built environment attributes impact household bicycle ownership or whether people with active lifestyle preferences involving bicycling (among other physically active recreational activities) deliberately choose to locate in neighborhoods that are conducive to bicycling enthusiasts. If such residential self-selection is taking place, then the true causal impacts of the built environment on bicycle ownership may not be as high as depicted in single equation models that treat residential location attributes as exogenous variables. Using a sample of more than 5000 households from the San Francisco Bay Area, a joint model of residential location choice and household bicycle ownership that accounts for self-selection effects and unobserved heterogeneity in the jointness is estimated and presented in this paper. The model results show that residential self-selection effects and heterogeneity in such effects can be substantial and ignoring these aspects of behavior may result in erroneous predictions of the true impacts of the built environment on household bicycle ownership.

Keywords: built environment, bicycle ownership, simultaneous equations model, residential self-selection, unobserved heterogeneity, modeling cause-and-effect, neighborhood type

1. INTRODUCTION

The use of non-motorized modes of transportation, notably walking and bicycling, for undertaking personal travel is an issue of considerable interest to the transportation planning profession. This interest stems from several key motivations. First, the use of non-motorized modes for personal travel is energy efficient and environmentally sustainable. Using such modes does not entail the consumption of fossil fuels and does not involve spewing toxic fumes into the earth's atmosphere that are typically associated with motor vehicle usage. Second, the use of non-motorized modes for personal travel constitutes a physically active transportation choice that has positive impacts on public health. Walking and bicycling, for any purpose, can help fight obesity and other public health challenges (e.g., cardiac disease) and as such, both transportation and public health officials are interested in promoting the use of these non-motorized modes. Third, providing safe walking and bicycling routes for users of these modes is a key consideration for transportation professionals. Hence, from a user safety standpoint, the profession is interested in being able to estimate demand for walking and bicycling accurately. Unfortunately, most traditional travel demand models do not adequately address demand estimation for these modes of transportation. While there may be additional motivations, these three points are sufficient to clearly point to the need for research and empirical studies in the use of non-motorized modes of transportation.

Within the context of analyzing the use of non-motorized modes of transportation, this paper focuses on bicycling. Although it would be ideal to analyze bicycle use (e.g., miles covered by bicycle, percent of trips by bicycle, etc.), such measures of bicycle use are often not well documented in travel surveys. Bicycle trips, particularly those for short recreational purposes, are subject to considerable under-reporting and the reporting of bicycle trip lengths, even where available, may be prone to error. Therefore, in this paper, total household bicycle ownership is used as a reasonable surrogate of household bicycle use; bicycle ownership has consistently been found to be a statistically significant determinant of bicycle usage (e.g., Cervero and Duncan, 2003; Simma and Axhausen, 2002; Cervero et al., 2002). It is possible that bicycle use and bicycle ownership are related in a circular or bi-directional relationship where, not only does bicycle ownership affect bicycle use, but bicycle use (or the preference to use the bicycle) affects bicycle ownership. Nevertheless, bicycle ownership can represent and determine the overall bicycle use for activities and travel, and capture the bicycling preferences of a household, to a substantial extent. Another reason for choosing bicycle ownership as the measure of interest in this paper is that it appears to be an under-studied variable. While there is some literature on bicycle trip making and trip attributes, there is very little analysis of bicycle ownership per se.

As mentioned earlier, the profession is interested in promoting the use of non-motorized modes of transportation. In the context of bicycling, land use – transportation planners and decision-makers are considering a range of land use – transportation policies and infrastructure configurations that would be potentially conducive to bicycling. These include higher density, mixed land use developments (thus facilitating the use of bicycle as many destinations are now located close to a person's residence and/or workplace), provision of a network of bicycle lanes/paths, and specific traffic safety measures that target bicycle users (exclusive signal indications at intersections, artificially lowered speed limits, etc.). With regard to the first item noted, i.e., higher density mixed land use development conducive to non-motorized transportation use, there is considerable interest in understanding the extent to which such built environment attributes can indeed impact bicycle use, or in the context of this paper, bicycle ownership. This is the central question addressed by this paper – what is the true causal impact of the residential built environment on bicycle ownership (and therefore, use)?

This question becomes complicated because the cause-and-effect relationship may not be a very clear one. While one may hypothesize that built environment attributes impact household bicycle ownership, it is also possible that the association is not causal, but simply associative. When treating residential built environment attributes as exogenous variables in a model of household bicycle ownership, one is assuming that the residential built environment is a given and ignoring the fact that residential built environment attributes are actually a manifestation of the residential location choice process exercised by households. In other words, residential location choice is endogenous to the choice phenomena under study; households with certain active lifestyle preferences may deliberately choose to live in neighborhoods that have land use configurations and transport infrastructure elements conducive to bicycling. If such residential self-selection effects are ignored, one can erroneously over-predict the impacts of land use – transport policies on bicycle ownership (and use). Is the relationship between the built environment and bicycle ownership completely causal or only purely associative? The truth probably lies somewhere in the middle; this paper is aimed at developing a model system that would directly contribute to answering this key question.

This paper makes a three-fold contribution to the literature. First, it sheds light on household bicycle ownership, a choice dimension that has hitherto been rarely studied and documented in the literature. Second, it involves the development of a joint model of residential location choice and household bicycle ownership that explicitly recognizes the self-selection phenomenon described in the previous paragraph. In the joint simultaneous equations model, error covariances represent the potentially common unobserved factors that impact both residential location choice and bicycle ownership (for example, desire to lead an athletic and active lifestyle) and serve as key indicators regarding the extent to which residential self-selection may be taking place. Third, the joint model is further enhanced to account for heterogeneity in residential self-selection effects and help determine the extent of simultaneity in decision-making with respect to these two choice phenomena. For example, although each household (or individual) may have its own life style preferences and corresponding residential self-selection preferences, low income households may face financial deterrents and other constraints (such as housing availability/affordability, market conditions, etc.) to self-select more into neighborhoods of their choice, when compared to higher income households. In another example, households with children may have a higher magnitude of residential self-selection preferences (effects) when compared to households without children, because of their desire to provide children with a family-oriented residential environment. The heterogeneity in the jointness, represented by the heterogeneity in the error covariances, captures such variation among households in residential self-selection effects. In summary, this is a unique study in the land use – travel behavior arena that presents a comprehensive analysis of the impact of socio-demographics and neighborhood characteristics on bicycle ownership while accounting for residential self-selection and heterogeneity in such effects.

Another substantive contribution of this paper is the use of factor analysis and clustering techniques to define a binary residential location variable that distinguishes the residential locations (i.e., the Traffic Analysis Zones) into bicycle-friendly and less bicycle-friendly neighborhoods. This binary variable is used to represent (as a dependent variable) the residential location choice component of the above mentioned joint (or heterogeneously-joint) model. The impact of the built environment on bicycle ownership levels, and the associated residential self-selection effects and corresponding heterogeneity are captured within the context of this binary variable (i.e., in the context of the impact of bicycle-friendly neighborhoods on bicycle ownership levels). Several studies in the past have used factor analysis and clustering techniques to distinguish residential locations into traditional/suburban/less bicycle-friendly and neo-traditional/bicycle-friendly neighborhoods; this study uses a combination of both techniques to come up with a bicycle-friendly neighborhood definition.

This paper is organized as follows. Following a brief literature review, a description of the data set and the methodology for defining the neighborhood type with respect to its bicycle-friendly character are provided. Then the model formulation is presented. Model estimation results and key conclusions are presented in the final sections.

2. URBAN FORM AND NON-MOTORIZED TRAVEL

There is a reasonably rich body of literature devoted to non-motorized travel demand analysis and it would be impossible to provide a comprehensive review of this literature within the scope of this paper. However, even a cursory review of the literature illustrates the level of interest and attention that has been and is being accorded to non-motorized transportation and bicycling in particular (e.g., Rietveld, 2001). Several measures of non-motorized travel and bicycle use have been analyzed in the past. These include such measures as trip rates, trip lengths and mileage, and mode choice/split for non-motorized travel. Baltes (1996) uses census data to identify factors influencing the choice of bicycle for work trips (non-discretionary trips). He finds that urban densities that promote shorter trips, relatively temperate climates, and a large proportion of students (e.g., university towns and communities) positively impact bicycle mode shares. Ewing et al. (2005) examine the choice of walking and bicycling to school for students based on school location and market area. With larger schools drawing student populations from ever-increasing market areas, they find that the extent to which students bicycle and walk has reduced over the years. Beck and Immers (1994) analyzed survey data collected in The Netherlands to identify reasons for choosing and not choosing to commute by bicycle. In most cases, the three main reasons for choosing to bicycle were speed, independence from public transit (and associated flexibility), and health advantages. Primary reasons reported for not commuting by bicycle include trip distance, discomfort, inability to travel with other people, and difficulty associated with transporting cargo or bags. Another study that examines bicycle use is that by Simma and Axhausen (2002) who report that bicycle ownership is strongly correlated with bicycle use. While this may appear to be a trivial finding at first glance, this is a key relationship that provides credence to the use of bicycle ownership as the dependent variable in this particular study. If it were found that bicycle ownership and use were poorly correlated (for example, households purchase and own bicycles, but never use them), then the potential use of bicycle ownership as a surrogate for bicycle use would have been questionable. There are a few studies that explicitly focus on bicycle ownership and use (e.g., Wigan, 1984; RTA, 2007).

There have been numerous studies that have examined the relationship between built environment and mode choice (including non-motorized modes). Cervero and Duncan (2003) use household activity data from the San Francisco region to study the links between urban environments and non-motorized travel. Their results reveal that areas with large city blocks are not pedestrian/bicycle friendly environments. The likelihood of bicycling increases with the number of bicycles in the household (just as studies show that driving increases with car ownership), although the possibility that this relationship is circular cannot be ignored – i.e., a desire to bicycle no doubt increases bicycle ownership. Among built environment features, Cervero and Duncan find that urban design and land use diversity factors are positively associated with the decision to ride a bicycle. Block size, grid street patterns, mixed land uses, jobs-housing balance, and availability of retail services within close proximity of one's origin generally encourage individuals to bicycle. Cervero et al. (2002) examine the impact of the City of San Francisco Car Share program on short-term travel behavior. They find that bicycle ownership, along with having transit passes, speedy transit services, low car ownership and participation in the City Car Share program, is closely associated with reduced private automobile usage. Rajamani et al. (2003) investigate the significance and explanatory power of a variety of urban form measures on non-work travel mode choice after controlling for demographic and level-of-service effects. They use the 1995 Portland (Oregon) Metropolitan Activity-Travel Survey data set for their analysis. They find that improvements in bicycle/pedestrian accessibility lead to increases in mode share for these modes in the context of recreational trips. Thus, providing a safe and comfortable pedestrian/bicycle environment, along with bringing shopping and recreational activity sites closer to residential neighborhoods, may be an effective way to increase bicycling and walking. However, one finding reported in the paper is the high cross-elasticity between walk and bicycle modes in relation to accessibility. Improvement in accessibility for one mode draws most share away from the other non-motorized mode. This suggests that it is important to improve accessibility for both modes simultaneously to increase non-motorized mode share as a whole. There are several other studies that examine the impact of urban measures on mode choice; they are not reviewed in detail in the interest of brevity. Example of such studies include the analysis of the built environment on non-motorized travel, including bicycle use (e.g., Cao et al., 2006; Chatman, 2005; Kitamura et al., 1997; Handy et al., 2006; Hunt and Abraham, 2007), miles traveled by various modes (e.g., Schwanen and Mokhtarian, 2005; Bagley and Mokhtarian, 2002), and mode choice (Pinjari et al., 2007). All of these papers generally report significant impacts of built environment attributes on mode choice and bicycle use; in addition, most note the potential effects of attitudes and values, lifestyle preferences, and residential self-selection that may play a role in shaping these impacts. In particular, Bhat and Guo (2007) and Cao et al. (2006) provide a detailed explanation of the notion of and review of studies addressing residential self-selection in models of travel behavior.

This paper is aimed at adding substantively to the body of literature on the nature of the true impacts of built environment variables on mode choice/use with particular focus on bicycle ownership (and, by association, bicycle use). The emphasis of this paper is on the formulation of a model system that accounts for residential self-selection effects and heterogeneity in these effects.

3. DATA

This section provides a brief description of the data used in the study.

3.1 Data Sources

The data used for this analysis is drawn from the 2000 San Francisco Bay Area Household Travel Survey (BATS) designed and administered by MORPACE International Inc. for the Bay Area Metropolitan Transportation Commission (MTC). This comprehensive activity-travel survey collected detailed socio-economic, demographic, and activity-travel information for a sample of about 15000 households in the Bay Area. Of particular interest to this study is that information about household vehicle and bicycle ownership, and residential locations was collected.

In addition to the 2000 BATS data, several other secondary data sources were used to derive spatial variables characterizing the activity-travel and built environment in the region. These include: (1) Zonal-level land-use/demographic coverage data, obtained from the MTC, (2) GIS layers of businesses (automotive businesses, shopping and grocery stores, medical facilities and personal services, food stores, sports and fitness centers, parks and gardens, restaurants, recreational businesses, and schools), obtained from the InfoUSA business directory, (3) GIS layers of bicycling facilities, also obtained from MTC, and (4) GIS layers of highway (interstate, national, state and county highways) network and local roadways (local, neighborhood, and rural roads) network, extracted from the Census 2000 Tiger files. From these secondary data sources, the following built environment variables were extracted and/or computed for the purpose of dividing the residential locations into bicycle-friendly and less bicycle-friendly neighborhoods:

1. Zonal size and density measures, such as total population, population density, household density, density of employment by each of several employment categories, and dummy variables for central business district (CBD), urban, suburban, and rural areas (computed based on employment density). These attributes were obtained from the zonal land-use data file.
2. Zonal land-use structure variables, such as housing type measures (fractions of single family and multiple family dwelling units), fractions of zonal area in residential and commercial land-uses, and land-use mix. The zonal land-use structure variables were constructed from the zonal land-use data file.
3. Zonal activity opportunity variables, such as activity center intensity (i.e., the number of business establishments per square mile) and density (i.e., the number of business establishments per square mile) for each of the following activity types (extracted from the InfoUSA business establishments data): (a) physically active recreation (fitness centers, sports centers, dance and yoga studios, parks, gardens, etc.), and (b) Natural recreation centers (parks, gardens, etc.)
4. Zonal transportation network measures, such as highway density (miles of highway facilities per square mile), bikeway density (miles of bikeway facilities per square mile), accessibility to other zones by bicycle mode, and local roadway density (miles of roadway density per square mile). These variables were extracted from the GIS layers of bikeways and roadways.

3.2 Definition of the Residential Neighborhood Type

The San Francisco Bay Area consists of 9 counties and 1099 Traffic Analysis Zones (TAZs) in all. This study uses factor analysis and clustering techniques to categorize the TAZs of the San Francisco Bay Area into bicycle-friendly and less bicycle friendly neighborhoods.

Several studies in the past have used either the clustering technique (Song and Knap, 2004; Rodriguez et al., 2005) and or the factor analysis method (Cervero, 1989; Cervero and Kockelman, 1997; Ewing et al., 2002; Srinivasan, 2002) to categorize residential locations into walking/bicycling friendly neighborhoods. In this context, it is important to note that a multitude of zonal land-use characteristics define the built environment and the bicycle-friendliness of a zone. The attributes include bicycling facilities (zonal bicycle lane density, length of bicycle lanes in the zone), bicycle route network (such as the number of zones accessible by the bicycle route network), other characteristics that may encourage bicycling (for example, the number of natural and physically active recreation centers in the zone), zonal density characteristics (zonal employment, population, and household densities), and the land-use structure (fraction of area under residential, commercial and other land uses, and the land-use mix). Also, many of these attributes may be very significantly correlated to each other (Cervero and Duncan, 2003). Thus, a combination of both the techniques should be used to come up with the definition of a bicycle-friendly neighborhood. Factor analysis helps in reducing the data (i.e., the various correlated attributes or *factors*) into a manageable number of *principal components* (or variables) that define the built environment of a neighborhood, and the clustering technique helps in using these *principal components* to divide the zones into bicycle-friendly and less bicycle-friendly neighborhoods.

Table 1 shows the results of the factor analysis (in the first block of the table) and cluster analysis (in the second block of the table) carried out for the San Francisco Bay Area. The six built environment characteristics (or *factors*) listed in the first column of the table were reduced to two *principal components* using the factor analysis. The factor loadings of the first *component* (in the second column) indicate that this *component* represents the residential density and land-use and the bicycling facilities in a zone. Thus, if a zone exhibits a high value of this *component*, that zone can be labeled as a residential type of zone with good bicycling facilities. Similarly, the second *component* captures zonal characteristics such as number of physically active centers such as sports centers, gymnasiums, and playing fields, etc., and number of natural recreation centers such as parks and gardens that can potentially encourage bicycling. The non-negligible loading (0.357) of the factor "bicycle lane density" on this component supports the notion that such activity centers may be associated with good bicycle facilities. Thus both the *components* represent bicycle-friendliness. The summary statistics indicate that the two *components* exhibit Thurstone's "deep structure" with eigen values above 1, and account for 67% of the variability in the six *factors* listed in the table.

After extracting the above mentioned two *components* from the factor analysis, a two-step cluster analysis is employed to divide the 1099 zones of the San Francisco Bay Area into two clusters, based on the two *components*. Subsequently, a descriptive analysis (for all the 1099 TAZs) was undertaken to analyze the zonal land use and bicycle facility characteristics (i.e., the *factors* used in the factor analysis) in the two clusters. Table 1 (in the second block) shows the average values of the zonal (or neighborhood) characteristics for the two clusters. Based on these values, the zones belonging to the cluster for which the average values of the *factors* are higher are labeled as bicycle-friendly neighborhoods and the zones belonging to the other cluster are labeled as less bicycle friendly neighborhoods. As can be seen, the bicycle friendly neighborhoods are characterized by better bicycling facilities, better accessibility by bicycle, higher density (street block density, and population density), and a larger number of physically active and natural recreational facilities. The fraction of residential land use was not substantially different across the two clusters. Overall, the neighborhood type definition based on a combination of factor analysis and cluster analysis appears to be intuitive and reasonable. This definition was used as a binary residential neighborhood type choice variable in the estimation of the heterogeneous-joint model. Of the 1099 TAZs, 320 were characterized as bicycle-friendly neighborhoods while the remaining were characterized as less bicycle-friendly neighborhoods.

3.3 Estimation Sample

The final estimation sample includes 5147 households from 5 counties (San Francisco, San Mateo, Santa Clara, Alameda, and Contra Costa) of the Bay area. The average bicycle ownership in these households is about 1.42 bicycles per household. Out of the 5147 households, 36.8% the households did not own bicycles, 22.5% owned one bicycle, 20.9% owned two bicycles, 8.8% owned three, 6.8% owned four, and 4.2% owned five or more bicycles. A descriptive analysis of the residential neighborhood type of these households indicates that 33.6% of the households reside in bicycle-friendly neighborhoods, while the remaining 66.4% of the households reside in less bicycle-friendly/suburban neighborhoods. A more extensive descriptive analysis of the sample is not included in this paper for the sake of brevity. The reader can find such information in several other sources (for example, see Morpace International Inc, 2002).

4. ECONOMETRIC MODELING FRAMEWORK

This section presents the econometric modeling methodology and framework employed in this study.

4.1 Model Structure

Let q (q = 1, 2, …, Q) be an index to represent households, and k (k = 1, 2, 3, …, K) be an index to represent bicycle ownership. r_q represents the residential neighborhood type chosen by household q; r_q = 1 if household q chooses a bicycle-friendly neighborhood and r_q = 0 if household q chooses a less bicycle-friendly neighborhood. Using these notational preliminaries, the structure of the residential neighborhood type choice model component is discussed first, the bicycle ownership model component is discussed second, the joint nature of the two components is discussed third, and the heterogeneity in the jointness of the two components is discussed at the end of this subsection.

4.1.1 The Residential Neighborhood Type Choice Component

The residential neighborhood type choice component takes the familiar binary logit formulation, as presented below, with r_q as the dependent variable:

$$u_q^* = (\beta' + \gamma_q')x_q + \eta_q + \varepsilon_q,\ r_q = 1 \text{ if } r_q^* > 0;\ r_q = 0 \text{ otherwise} \tag{1}$$

In the equation above, u_q^* is the indirect utility that household q obtains from locating in a bicycle-friendly residential neighborhood, x_q is an (M x 1)-column vector of socio-demographic attributes (including a constant) associated with household q (for example, household size, income, housing type, etc.). β represents a corresponding (M x 1)-column vector of mean effects of the elements of x_q on the utility associated with neighborhood choice, while γ_q is another (M x 1)- column vector with its m^{th} element representing unobserved factors specific to household q that moderate the influence of the corresponding m^{th} element of the vector x_q. η_q captures common unobserved factors influencing household q's utility for a neo-traditional/bicycle-friendly neighborhood type choice and the household's bicycle ownership propensity (more details on this later in this subsection). ε_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across individuals q.

4.1.2 The Bicycle Ownership Model Component

The household bicycle ownership component takes the ordered logit formulation, as presented below:

$$y_q^* = (\alpha' + \delta_q')z_q \pm \eta_q + (\theta + \mu' w_q + \lambda_q)r_q + \xi_q,\ y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k \tag{2}$$

In the equation above, y_q^* is the latent propensity associated with the bicycle ownership of household q. This latent propensity y_q^* is mapped to the actual bicycle ownership level y_q (i.e., the number of bicycles owned by the household) by the ψ thresholds ($\psi_0 = -\infty$ and $\psi_K = \infty$) in the usual ordered-response fashion. z_q is an (L x 1) column vector of attributes (not including a constant and not including the household's residential neighborhood type) that influences the propensity associated with bicycle ownership. α is a corresponding (L x 1)-column vector of mean effects, and δ_q is another (L x 1)-column vector of unobserved factors moderating the influence of attributes in z_q on the bicycle ownership propensity for household q. As discussed in the previous section, η_q captures common unobserved factors influencing household q's utility for a neo-traditional/bicycle-friendly neighborhood type choice and the household's bicycle ownership propensity. θ is a scalar constant representing the effect of residential neighborhood type (i.e., r_q) on household bicycle ownership, w_q is a set of household attributes that moderate the effect of residential neighborhood type on household bicycle ownership, and μ is a corresponding vector of coefficients. λ_q is an unobserved component influencing the impact of residential neighborhood type for household q, and ξ_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across households q.

4.1.3 The Joint Model System

The joint nature of the model system arises because of the presence of the common unobserved η_q term in the residential neighborhood type choice (see Equation 1) and bicycle ownership (see Equation 2) model components. More generally, the model system allows self-selection of households (based on their bicycle ownership preferences) into neighborhoods based on unobserved preferences and other unobserved factors[1]. For example, households with physical-fitness oriented and bicycling oriented individuals are likely to associate a higher utility (than that of other observationally identical households) to residing in bicycle-friendly neighborhoods, as well as a higher prosperity to own a higher number of bicycles. Thus such households may self select to reside in bicycle-friendly neighborhoods and have a higher bicycle ownership.

The '±' sign in front of the η_q term in the bicycle ownership propensity equation indicates that the correlation in the unobserved factors may be positive or negative. If the sign is positive (negative), it implies that individuals who intrinsically have a higher (lower) inclination to reside in bicycle-friendly neighborhoods tend to have a higher bicycle ownership propensity. From an intuitive standpoint, we expect a positive correlation. However, one can empirically test the models with both '+' and '−' signs to determine the best empirical result. In any case, if the correlation due to common unobserved factors is ignored, when actually present, it results in an erroneous estimation of the impact of the residential neighborhood type on bicycle ownership. More specifically, if the unobserved correlation between bicycle-friendly neighborhood choice utility and bicycle ownership propensity is positive, as we expect, ignoring this correlation would

[1] It is to be noted that the model also allows for residential self-selection due to observed factors. Inclusion of observed factors such as sociodemographic variables that affect a household's residential neighborhood selection in its bicycle ownership component accounts for the self-selection of that household due to those observed factors.

result in an inflated impact of the bicycle-friendly residential neighborhood on household bicycle ownership.

To complete the model structure of the system in Equation (1), it is necessary to specify the structure for the unobserved vectors γ_q and δ_q, and the unobserved scalars λ_q and η_q. In the current paper, it is assumed that the γ_q and δ_q elements, and λ_q and η_q, are independent realizations from normal population distributions; $\gamma_{qm} \sim N(0, \upsilon_m^2)$, $\delta_{ql} \sim N(0, \omega_l^2)$, $\lambda_q \sim N(0, \tau^2)$, and $\eta_q \sim N(0, \sigma^2)$.

4.1.4 The Heterogeneous-Joint Model System

The joint nature of the model system may be allowed to vary across households by allowing the magnitude of the common unobserved factors to vary based on household characteristics. That is, the standard deviation of the common unobserved factors (i.e., the standard deviation of the η_q term) can be expressed as a function of household characteristics as: $\sigma = \exp(\iota + \varpi' v)$, where ι is a constant, v is a vector of household characteristics, and ϖ is the corresponding coefficient vector. The common unobserved η_q term in Equations (1) and (2) can then be expressed as $\eta_q = \vartheta_q \exp(\iota + \varpi' v)$, where $\vartheta_q \sim N(0,1)$. Thus, the joint model system accounting for unobserved heterogeneity in residential self-selection effects can be expressed as:

$$u_q^* = (\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v) + \varepsilon_q,\ r_q = 1 \text{ if } r_q^* > 0;\ r_q = 0 \text{ otherwise,}$$
$$y_q^* = (\alpha' + \delta_q')z_q \pm \vartheta_q \exp(\iota + \varpi' v) + (\theta + \mu' w_q + \lambda_q) r_q + \xi_q,\ y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k \qquad (3)$$

In the above heterogeneous-joint model formulation, the first equation represents the binary logit model component for the household's choice of residing in a bicycle-friendly neighborhood, while the second equation represents the ordered logit model component for household bicycle ownership. The common $\vartheta_q \exp(\iota + \varpi' v)$ term across both of the model components, which is a function of household characteristics, allows for the possibility that the residential self-selection effects due to common unobserved factors vary across households. For example, households with children may have a higher magnitude of residential self-selection preferences when compared to households without children, because of their desire to provide the children with a more family-oriented residential environment. Thus households with children may have a higher magnitude of common unobserved factors. This can be captured by specifying the number of children in the household as a variable in the v vector, with an expectation that the corresponding coefficient will be positive.

It is important to recognize that the heterogeneity in residential self-selection effects is a form of behavioral heterogeneity in the population; ignoring population heterogeneity of any form when it is present will generally lead to biased and inconsistent estimation, inferior model fit, loss of behavioral interpretations and distortion of policy implications. Hence it is important to, at the least, test for the presence of heterogeneity (across decision makers) in the covariance between the two models.

It is to be noted here that the heterogeneous-joint model formulation presented in this paper nests a homogenous-joint model with uniform homogeneous residential self-selection effects across the population (when $\sigma = \exp(t)$; i.e., when ϖ is a zero vector) and a disjoint model system with no residential self-selection effects (when $\sigma = 0$; i.e., when $t \to -\infty$, and ϖ is a zero vector). Thus one can use the same model formulation to statistically test for the presence of and heterogeneity in residential self-selection effects in the context of modeling bicycle ownership.

4.2 Model Estimation

The parameters to be estimated in the joint model system of Equation (1) are the β, α, μ, and ϖ vectors, the θ and ι scalars, the ψ thresholds, and the following variance terms: υ_m^2, ω_l^2, and τ^2. Let Ω represent a vector that includes all of these parameters to be estimated[2]. Also, let c_q be a vector that vertically stacks the γ_q and δ_q vectors, and the λ_q and ϑ_q scalars. Let Σ be another vertically stacked vector of standard errors υ_m, ω_l, and τ, and let $\Omega_{-\Sigma}$ represent a vector of all parameters except the standard error terms. Let d_{qk} be a dummy variable taking the value 1 if household q owns k number of bicycles and 0 otherwise. Finally, let G(.) be the cumulative distribution of the standard logistic distribution. Then, the likelihood function, for a given value of $\Omega_{-\Sigma}$ and error vector c_q, may be written for household q as:

$$
\begin{aligned}
&L_q(\Omega_{-\Sigma} \mid c_q) = \\
&\left\{ \frac{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right]}{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right] + 1} \right\}^{r_q} \left\{ \frac{1}{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right] + 1} \right\}^{1-r_q} \\
&\times \left\{ \begin{array}{l} G\left[\psi_k - \left\{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)r_q \pm \vartheta_q \exp(\iota + \varpi' v)\right\}\right] \\ -\, G\left[\psi_{k-1} - \left\{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)r_q \pm \vartheta_q \exp(\iota + \varpi' v)\right\}\right] \end{array} \right\}^{d_{qk}},
\end{aligned}
\tag{4}
$$

The unconditional likelihood function can be computed for housheold q as:

$$
L_q(\Omega) = \int_{c_q} (L_q(\Omega_{-\Sigma}) \mid c_q)\, dF(c_q \mid \Sigma), \tag{5}
$$

where F is the multidimensional cumulative normal distribution. The log-likelihood function for all the households can be written as: $L(\Omega) = \sum_q L_q(\Omega)$. Simulation techniques are applied to approximate the multidimensional integral in Equation (5), and the resulting simulated log-likelihood function is maximized. Specifically, the scrambled Halton sequence (see Bhat, 2003) is used to draw realizations from the population normal distribution. In the current paper, the sensitivity of parameter estimates was tested with different numbers of scrambled Halton draws per observation, and the results were found to be stable with as few as 100 draws.

---

[2] The reader will note here that the υ_m^2 terms of the binary logit model component can not be estimated because those parameters are not theoretically identifiable.

5. MODEL ESTIMATION RESULTS

This section presents a summary of the model estimation results together with key findings and behavioral interpretations that may be drawn from the models. A series of models were estimated, including:

- A heterogeneous-joint model system of residential location and bicycle ownership
- A homogenous-joint model system of residential location and bicycle ownership
- A disjoint (or independent) model system of residential location and bicycle ownership
- A disjoint (or independent) model system including only a constant in the residential location model and no explanatory variables in the bicycle ownership model

For the sake of brevity, only the first model listed above, i.e., the heterogeneous-joint model system, is presented in this paper in its entirety. Appropriate log-likelihood ratio tests are applied to test the significance of residential self-selection effects and unobserved heterogeneity by comparing the model systems listed above.

Model estimation results for the heterogeneous-joint model system are presented in Table 2. The first part of the table shows the binary logit model of residential location choice (bicycle-friendly neighborhood type choice = 1). The constant does not have a substantive interpretation and is statistically insignificant. Similarly, the age of the householder is statistically insignificant. The weak negative coefficient suggests that more mature households with older householders are less inclined to locate in bicycle-friendly neighborhoods. It is interesting to note that the number of children under 16 years of age, living in a single-family dwelling unit, and owning a house are all negatively associated with choosing to live in a bicycle-friendly neighborhood. This is consistent with expectations; all of these attributes are associated with living in suburban neighborhoods that typically have poorer bicycle-friendly attributes.

The ordered-response logit model of bicycle ownership is presented in the second block of Table 2. All of the explanatory variables included in the model are statistically significant. Bicycle ownership is positively associated with the number of active adults in the household, the number of children in the household, and the number of students in the household. However, single individuals and older households show a negative tendency towards owning bicycles. Where the householder is male, the household is Caucasian, and the household annual income is high, there is a tendency to own more bicycles. Similarly, residing in a single-family dwelling unit and owning a household are positively associated with bicycle ownership. This is an interesting finding in that single-family dwelling units and home ownership are more likely to be associated with traditional surburban neighborhoods that may not be as bicycle-friendly as neo-traditional urban neighborhoods. Moreover, the positive coefficients are statistically significant even after controlling for the number of active adults, number of children, and number of students in the household. In the San Francisco Bay Area, it is likely that the temperate climate and active lifestyle preferences contribute to higher levels of bicycle ownership even in traditional suburban neighborhoods. This is not to say that bicycle-friendly neighborhoods have no impact on bicycle ownership. Even after controlling for all other socio-economic and demographic variables, it is found that household location in a bicycle-friendly neighborhood significantly impacts bicycle ownership in a positive way.

The third block of the table shows coefficient estimates for variables in the standard deviation equation of the common error component between the residential location choice and bicycle ownership equations (the η_q term). Recall that these variable effects are representative of the heterogeneity in residential self-selection effects. It is found that heterogeneity in residential self-selection effects is primarily due to the presence of children, both young children less than 5 years of age and older children between 5 and 16 years of age. This finding is consistent with the hypothesis presented earlier that the presence of children may contribute to heterogeneity in residential self-selection as households with children choose to locate in neighborhoods that are potentially more family-oriented and conducive to raising children. As the number of children increases, the extent of heterogeneity increases. In addition, a modest impact of income on heterogeneity in residential self-selection effects is seen. Heterogeneity in residential self-selection is found to be greater for low income households making less than $35,000 per year. This is also consistent with the notion that low income households may be limited with respect to their ability to choose to locate in neighborhoods of their preference due to income limitations.

The final block of the table presents a comparison of log-likelihood measures for the four different model systems listed earlier in this section. A rather interesting finding from this table is that the homogenous-joint model (not reported in this paper) did not show the presence of statistically significant residential self-selection effects. That is, the standard deviation of the common error component (σ) was statistically insignificant even at a 20% level of significance, with the t-statistic of σ being close to zero. Also, as shown in the table, the homogenous-joint model did not show any significant improvement in the log likelihood when compared to the log likelihood of independent residential neighborhood type choice and bicycle ownership models. The heterogeneous-joint model, on the other hand, showed a statistically significant improvement in the log-likelihood. From the table, a nested log-likelihood ratio test between the heterogeneous-joint and homogenous-joint models (a log-likelihood ratio statistic = 31.36 with 3 degrees of freedom) indicates that there is significant variance in residential self-selection effects in the population.

In order to further assess the importance of capturing residential self-selection effects and corresponding heterogeneity, the t-statistic of the standard deviation of the common error component (σ) in the heterogeneous-joint model was calculated for different population segments such as households with children, households without children, households with low annual income, and households with high annual income. That is, the t-statistics for $\sigma = \exp(\iota + \varpi' v)$ were computed for different values of v. Since, the estimates and the t-statistics of ι and ϖ are known, the t-statistics of σ for each demographic segment could be computed in a straightforward manner applying the delta method. Details of the delta method are available in any standard econometrics text book such as Wooldridge (2002).

The t-statistics of σ for each segment were around 1.0, indicating the marginally significant presence of residential self-selection effects in each segment. These t-statistics of σ for each demographic segment in the heterogeneous-joint model are higher than that of the t-statistic of σ in the homogenous-joint model (which was close to zero). This is perhaps why there was a statistically significant improvement in the log-likelihood; the net effect of capturing differential residential self-selection effects in different demographic segments might have contributed to the improvement in the log-likelihood. This supports the notion that one needs to account for heterogeneity in self-selection effects, although, in the absence of significant residential self-selection effects, it is unclear whether accounting for any further heterogeneity in such effects offers significant advantages in a policy analysis context. A policy simulation analysis that assists in such an assessment is presented in the next section.

6. POLICY SIMULATION AND ANALYSIS

The coefficient estimates of the bicycle ownership component of the heterogeneous-joint model system can be applied to predict the changes (aggregate-level elasticities) in bicycle ownership due to changes in socio-demographic characteristics and built environment attributes. The impact of the built environment variables was computed in two ways: (1) by computing the elasticity effect of the binary residential neighborhood type variable, and (2) by computing the elasticity effect due to a 25% change in the built environment attributes used to define the residential neighborhood type variable.

The aggregate-level pseudo elasticity of an ordinal exogenous variable (such as the number of active adults in the household) was computed by increasing the value of the ordinal variable by unity for each household and obtaining the relative change in expected aggregate bicycle ownership level. Next, to compute an aggregate-level pseudo elasticity of a dummy variable (such as the neighborhood type), the value of the variable is changed to one for the subsample of observations (i.e., households) for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. Then the shifts in expected aggregate bicycle ownership level in the two subsamples are summed after reversing the sign of the shifts in the second subsample. This is an effective proportional change in expected aggregate bicycle ownership level in the entire sample due to a change in the dummy variable from 0 to 1. The aggregate-level "arc" elasticity of a continuous exogenous variable (such as household income) is computed by increasing the continuous variable by a uniform 25% across all households and obtaining the proportional change in the expected aggregate bicycle ownership level. For the built environment variables, the aggregate "arc" elasticity was computed in the following manner. First, the values of the built environment attributes for only the less bicycle-friendly neighborhoods was increased by 25%. Then the neighborhoods were re-categorized into a binary neighborhood type variable based on their new built environment characteristics. The subsequent changes in expected aggregate bicycle ownership are reported as the elasticity values.

Table 3 shows the aggregate level (pseudo/arc) elasticity effects on the expected bicycle ownership levels in the estimation sample for the explanatory variables in the bicycle ownership component of the heterogeneous-joint model. Several observations can be made from the table. First, the elasticity effects of the socio-demographic variables are higher than those of either the residential neighborhood type variable or the built environment attributes. Second, the elasticity effects of the built environment attributes are smaller compared to that of the neighborhood type variable. While the elasticity effect of the built environment variables represents the effect of addressing individual attributes of a neighborhood, the elasticity effect of the neighborhood type variable represents an overall effect of changes in all of the built environment variables. This indicates that built environment policies may be more effective when used in combination (i.e., for example, increase the bicycling facilities in the neighborhood, and also increase the connectivity of the bicycle route network to other neighborhoods) rather than modifying individual elements (such as increase only the bicycle facilities) of a neighborhood.

The aggregate elasticity effects reported in Table 3 were computed using the bicycle ownership component of the heterogeneous-joint modeling system. In order to assess the impact of the residential self-selection effects, the elasticity effects (for the neighborhood type and built environment variables) were computed using the estimates of a disjoint model system in which an independent bicycle ownership model was estimated with no residential self-selection effects. The elasticity effects of the independent bicycle ownership model are not shown in the table because they were not perceivably different from those of the heterogeneous-joint model. The elasticity effects were not different even for the different socio-demographic segments (such as households with children and households without children) for which the standard deviation of the common error component (i.e., the residential self-selection effects) was estimated to be marginally significant. This indicates that, in the current empirical context considered in this paper, residential self-selection does not appear to have a significant effect on the assessment of the impact of residential neighborhood type and built environment attributes on bicycle ownership in any segment of the population. Although there was a significant improvement in the log-likelihood due to capturing heterogeneity in residential self-selection effects, the effect of self-selection is itself not significant enough to gain much advantage from capturing any further heterogeneity. This may be because the residential self-selection preferences are already captured in the bicycle ownership model by including socio-demographic variables. In other words, in the current context, there are no significant self-selection effects due to unobserved factors in the assessment of the effect of variables. However, it is important, at the least, to test for the presence of residential self-selection and the heterogeneity in self-selection before using a simple ordered logit model for bicycle ownership.

7. CONCLUSIONS

Bicycle ownership and use is of much interest to the land use and transportation planning profession that is interested in promoting safety, energy efficient and environmentally sustainable transportation mode use, and healthy lifestyles. However, there is the classic debate as to whether built environment attributes (or residential neighborhood type) significantly impact bicycle ownership and use due to the residential self-selection effects that may be at play. People with a proclivity towards healthy and active lifestyles or bicycle use may self-select themselves to locate in neighborhoods that are conducive to bicycling. In other words, it is lifestyle preferences that affect bicycle ownership and use as opposed to the built environment attributes themselves. Treating built environment attributes as exogenous variables in models of bicycle ownership and use (or any travel behavior model) may lead to grossly inflated and erroneous estimates of changes in behavior in response to changes in policy or land use – transport systems design configurations.

In order to shed light on this phenomenon in the context of bicycle ownership, a rather under-studied aspect of mobility in the transportation research arena, this paper presents a joint model of residential location and bicycle ownership that accounts for residential self-selection effects and accommodates unobserved heterogeneity in such effects. Using a sample of about 5000 households from the San Francisco Bay Area, a series of joint models with and without residential self-selection effects, and with and without unobserved heterogeneity, are estimated using simulated maximum likelihood estimation approaches. A hybrid factor- and cluster-analysis approach is used to group residential locations (denoted by traffic analysis zones) as being either bicycle-friendly or less bicycle-friendly. Residential location choice is then treated as a binary choice while bicycle ownership is treated as an ordered response variable. Model estimation results provide intuitively meaningful results with a host of demographic and socio-economic variables affecting residential location choice and bicycle ownership. More importantly, it is found that, even after controlling for all other factors, neighborhood type significantly impacts bicycle ownership. Thus, it appears that policies aimed at modifying the built environment attributes could significantly affect bicycle ownership (and therefore, bicycle use as well). However, it is interesting to note that the effect of residential neighborhood type on bicycle ownership is generally lower than that for other socio-economic and demographic variables such as income, number of children, age, household composition, and type of housing unit.

In the current empirical context, it appears that the observed variables included in the model may have captured many of the self-selection effects and it is questionable whether additional benefits would be obtained by accounting for heterogeneity in residential self-selection effects when such effects themselves are not very statistically significant. Indeed, this is further illustrated by the policy simulation experiment presented in this paper; elasticity estimates are virtually unchanged when one considers the heterogeneous-joint model system against the independent (disjoint) model system that includes no residential self-selection effects whatsoever. The policy analysis further suggests that strategies aimed at enhancing bicycle ownership (use) are best implemented as a package that includes attention to bicycle lane density, bicycle accessibility to destinations, street block density, and availability of recreational centers around the residential neighborhood. The policy simulation suggests that implementing strategies in isolation may yield little benefits.

Future research in this arena should focus on examining the consistency of this finding across multiple geographic contexts and associated data sets. In addition, the spatial unit of analysis used in the context of residential location choice modeling (the traffic analysis zone is used in this paper) merits further consideration to examine whether the findings are robust and consistent even if different levels of spatial resolution are used to model residential location choice.

REFERENCES

Bagley, M. N. and P.L. Mokhtarian (2002) The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. *Annals of Regional Science*, 36(2), 2002, pp. 279-297.

Baltes M.R. (1996) Factors Influencing Nondiscretionary Work Trips by Bicycle Determined from 1990 U.S. Census Metropolitan Statistical Area Data, *Transportation Research Record* 1538, pp. 96-101.

Beck, M.J.H. and L. H. Immers (1994) Bicycle Ownership and Use in Amsterdam. *Transportation Research Record* 1441, pp. 141-148.

Bhat, C. R. (2003) Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research*, 37B(9), pp. 837-855.

Bhat, C.R. and J.Y. Guo (2007) A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. *Transportation Research Part B*, 41(5), pp. 506-526.

Cao, X., S.L. Handy, and P.L. Mokhtarian (2006) The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX. *Transportation*, 33(1), pp. 1-20.

Cervero, R. (1989) *America's Suburban Centers: The Land Use-Transportation Link*. Unwin-Hyman, Boston.

Cervero. R., N. Creedman, M. Pohan, and M. Pai (2002) City CarShare: Assessment of Short-Term Travel-Behavior Impacts. Prepared for Department of Transportation & Parking, City of San Francisco, CA.

Cervero R. and M. Duncan (2003) Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. Final Summary Report, University of California Transportation Center, University of California, Berkeley.

Cervero, R. and K. Kockelman (1997), Travel Demand and the 3Ds: Density, Diversity and Design. *Transportation Research Part D,* 2, pp. 199-219.

Chatman, D.G. (2005). How the Built Environment Influences Non-work Travel: Theoretical and Empirical Essays. Ph.D. Dissertation, Urban Planning, University of California, Los Angeles.

Ewing, R., R. Pendall, and D. Chen (2002) *Measuring Sprawl and Its Impacts.* Smart Growth America, Washington, D.C.

Ewing, R., W. Schroeer, and W. Greene (2005), School Location and Student Travel Analysis of Factors Affecting Mode Choice, *Transportation Research Record* 1895, pp. 55-63.

Handy, S., X. Cao, and P.L., Mokhtarian (2006). Self-Selection in the Relationship Between Built Environment and Walking? Evidence from Northern California. *Journal of the American Planning Association,* 72(1), pp. 55-74.

Hunt, J.D., and J.E. Abraham (2007) Influences on Bicycle Use. *Transportation,* 34(4), pp. 453-470.

Kitamura, R., P.L. Mokhtarian, and L. Laidet (1997) A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. *Transportation,* 24(2), pp. 125-158.

MORPACE International, Inc. (2002) Bay Area Travel Survey Final Report, March 2002. ftp://ftp.abag.ca.gov/pub/mtc/planning/BATS/BATS2000/

Pinjari, A.R., R.M. Pendyala, C.R. Bhat, and P.A. Waddell (2007b) Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice. Forthcoming in *Transportation*.
Rajamani, J., C.R. Bhat, S. Handy, G. Knaap, and Y. Song (2003) Assessing the Impact of Urban Form Measures in Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects, *Transportation Research Record* 1831, pp. 158-165.
Rietveld, P. (2001), Biking and Walking: The Position of Non-Motorized Transport Modes in Transport Systems. In K. J. Button and D.A. Hensher (Ed.), *Handbook of Transport Systems and Traffic Control,* pp. 299-319, Pergamon, London, United Kingdom.
Roads and Traffic Authority (RTA) (2007) Cycling in Sydney: Bicycle Ownership and Use. Prepared by the Transport Data Centre, New South Wales Department of Transport.
Rodriguez, D.A, H.M. Young, and R.J. Schneider (2005) An Easy to Compute Index for Identifying Built Environments that Support Walking. Presented at the 85th Annual Meeting of the Transportation Research Board, January, Washington, D.C.
Schwanen, T. and P.L. Mokhtarian (2005) What If You Live in the Wrong Neighborhood? The Impact of Residential Neighborhood Type Dissonance on Distance Traveled. *Transportation Research Part D,* 10(2), pp. 127-151.
Simma, A. and K.W. Axhausen (2001) Structures of Commitment in Mode Use: A Comparison of Switzerland, Germany and Great Britain, Conference paper STRC 2001. First Swiss Transport Research Conference Monte Verità/Ascona, March 2001.
Song, Y. and G. J. Knapp (2004) Measuring Urban Form, *Journal of the American Planning Association,* 70, pp. 210-226.
Srinivasan, S. (2002) Quantifying Spatial Characteristics of Cities. *Urban Studies,* 39, pp. 2005-2028.
Wigan, M.R. (1984) Bicycle Ownership, Use and Exposure: Participation and Activity Patterns in Melbourne, Australia. *Transportation Research Part A*, 18(5-6), pp. 379-398.
Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

List of Tables

TABLE 1 Results of Factor Analysis and Cluster Analysis

Factor Analysis Results: Factor Loadings and Summary*

Components** \ Factors	Bicycle facilities, residential density and land-use	Activity centers that encourage bicycling
Bicycle lane density (mileage per square mile)	0.513	0.357
Number of zones accessible from the home zone by bicycle	0.784	
Street block density (mileage per square mile)	0.924	
Household population density (per acre)	0.839	
Fraction of residential land use in the zone	0.716	
Number of physically active and natural recreation centers in the zone		0.914
Summary statistics		
Eigen value	2.95	1.06
Percentage of variance accounted by the component	49.12	17.71

Cluster Analysis Results: Zonal-level Land Use Characteristics (Averages) by Neighborhood Type

Characteristic \ Neighborhood Type	Bicycle-friendly neighborhood	Less bicycle-friendly neighborhood
Bicycle lane density (mileage per square mile)	4.92	1.88
Number of zones accessible from the home zone by bicycle	60.97	29.31
Street block density (mileage per square mile)	21.00	13.92
Household population density (per acre)	20.70	7.73
Fraction of residential land use in the zone	0.56	0.49
Number of physically active and natural recreational centers in the zone	5.23	1.16

* Principal components estimation and varimax rotation were used in deriving the results
**Factor loadings below 0.35 below are considered insignificant and not shown in the table

TABLE 2 Estimation Results of the Heterogeneous-Joint Residential Neighborhood Choice and Bicycle Ownership Choice Model

Variables in the residential neighborhood choice (binary logit) component[a]	Parameter	t-stat
Constant	0.1146	0.80
Age of the householder	-0.0033	-1.00
Number of children (of age < 16 years) in the household	-0.1431	-2.91
Household lives in a single family dwelling unit	-0.6030	-6.78
Own house	-0.6224	-6.71
Variables in the bicycle ownership choice (ordered response) component		
Number of active adults in the household	0.3043	5.53
Number of children (of age < 5 years) in the household	0.4224	6.49
Number of children (of age between 5 and 16) in the household	1.0691	15.08
Number of students in the household	0.3220	5.70
Single person household	-0.3047	-3.31
Age of householder greater than 60 years	-0.6381	-6.37
Householder is male	0.1248	2.38
Caucasian household	0.5977	9.69
Household annual Income in 10,000s of dollars	0.4500	7.53
Household lives in a single family dwelling unit	0.3962	5.73
Own household	0.2788	4.03
Household location in a neo-traditional/bicycle-friendly neighborhood	0.1794	2.96
Variables in the standard deviation equation of the common error component between residential neighborhood and bicycle ownership models		
Constant	-4.2668	-4.18
Number of children (of age < 5 years) in the household	0.7850	3.63
Number of children (of age between 5 and 16) in the household	1.3818	5.13
Household annual Income less than $35K	0.8230	1.04
Log-Likelihood Measures		
Model	**Log-likelihood**	**Number of parameters**
Heterogeneous-joint residential neighborhood choice and bicycle ownership choice model	-10259.55	27
Homogenous-joint residential neighborhood choice and bicycle ownership choice model	-10275.23	24
Independent residential neighborhood choice and bicycle ownership choice model	-10275.32	23
Independent models with only a constant as the explanatory variable in the binary logit model, and no explanatory variables in the ordered logit model	-11461.52	7

[a] The variables are in the utility equation of bicycle-friendly neighborhood type choice.

TABLE 3 Elasticity Effects of Variables in the Bicycle Ownership Component of the Heterogeneous-Joint Model

Sociodemographic variables in the bicycle ownership component of the heterogeneous-joint model	Elasticity Effect (%)
Number of active adults in the household	13.04
Number of children (of age < 5 years) in the household	17.95
Number of children (of age between 5 and 16) in the household	47.47
Number of students in the household	13.82
Single person household	-12.65
Age of householder greater than 60 years	-25.10
Householder is male	5.20
Caucasian household	24.10
Household annual Income in 10,000s of dollars	4.48
Household lives in a single family dwelling unit	16.53
Own household	11.57
Residential neighborhood type variable	7.50
Built Environment variables used to define the neighborhood type	
Bicycle lane density (mileage per square mile)	1.22
Number of zones accessible from the home zone by bicycle	1.27
Street block density (mileage per square mile)	1.42
Number of physically active and natural recreational centers in the zone	1.01