A JOINT MODEL OF RESIDENTIAL NEIGHBORHOOD TYPE CHOICE 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 neighborhood type choice and bicycle ownership. The objective is to isolate the true causal effects of the neighborhood attributes on household bicycle ownership from spurious association due to residential self-selection effects. The joint model accounts for residential self-selection due to both observed socio-demographic characteristics and unobserved preferences. In addition, the model allows for differential residential self-selection effects across different socio-demographic segments. The model is estimated using a sample of more than 5000 households from the San Francisco Bay Area. Further, a policy simulation analysis is carried out to estimate the impact of neighborhood characteristics and socio-demographics on bicycle ownership.

The model results show a substantial presence of residential self-selection effects due to observed socio-demographics such as number of children, dwelling type, and house ownership. It is shown for the first time in the self-selection literature that ignoring such observed self-selection effects may not always lead to overestimation of the impact of neighborhood attributes on travel related choices such as bicycle ownership. In the current context, ignoring self-selection due to socio-demographic attributes resulted in an underestimation of the impact of neighborhood attributes on bicycle ownership. In the context of unobserved factors, no significant self-selection effects were found. However, it is recommended to test for such effects as well as heterogeneity in such effects before concluding that there are no unobserved factors contributing to residential self-selection.

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

1. INTRODUCTION

1.1 Non-Motorized Travel and Bicycle Ownership

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. The key motivation behind this interest is that travel by non-motorized modes constitutes an environmentally sustainable and a physically active transportation choice, which both transportation and public health officials are interested in promoting. As a result of the interest in promoting non-motorized transportation, a number of research studies have attempted to analyze and identify the determinants of non-motorized travel demand. Even a cursory review of the literature illustrates the level of interest and attention accorded to analyzing non-motorized travel behavior [see, for example, (*1*-*6*)].

Within the context of non-motorized travel behavior and demand analysis, bicycle ownership appears to be a relatively understudied variable. While there is some literature on bicycle usage and such travel measures as trip rates, travel mileage, and mode choice [see, for example, (*3*), (*11*), and (*7*-*12*)] there is very little analysis of bicycle ownership per se [see, (*13*) and (*14*) for such sparsely available bicycle ownership studies). Thus, household level bicycle ownership is the focus of this study.

It is possible that bicycle use and bicycle ownership are related in a bi-directional relationship where, not only does bicycle ownership affect bicycle use, but bicycle use (or related preferences) affects bicycle ownership. Thus it would be ideal to analyze bicycle use along with bicycle ownership. However, measures of bicycle use (e.g., miles covered by bicycle, percent of trips by bicycle, etc.) are often not well documented and subject to under-reporting and inaccuracy in travel surveys [see (*15*)]. Nevertheless, bicycle ownership can be assumed to represent and determine the overall bicycle use for activities and travel, and capture the bicycling preferences of households and individuals. Further, bicycle ownership has been consistently found to be an important determinant of bicycle usage [see, for example, (*10*), and (*16*-*19*) for findings that indicate a statistically significant association of bicycle ownership with bicycle usage for several activities and/or related travel). Thus it is important to identify the socio-demographic, land-use, and transportation system characteristics that are positively associated with bicycle ownership levels.

1.2 The Residential Self-Selection Phenomenon

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 policies and infrastructure configurations that would be potentially conducive to bicycling. These include higher density mixed land use developments, walk/bicycle-friendly neighborhoods, and specific traffic safety measures that target bicycle users. With regard to the first two items noted, i.e., higher density mixed land use bicycle-friendly neighborhood development conducive to non-motorized transportation use, there is considerable interest in understanding the extent to which such neighborhoods 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 bicycle-friendly neighborhood 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 (such a bicycle-friendly nature) has a significant impact on household bicycle ownership, it is also possible that the association is not causal, but simply associative. When treating the residential built environment as exogenous to a model of household bicycle ownership, one is ignoring the possibility of the residential neighborhood choice process exercised by households. In other words, residential neighborhood 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 answering this key question [see Bhat and Guo (*20*), and Cao et al. (*21*) for a detailed explanation of the notion of residential self-selection, and thorough reviews of studies addressing residential self-selection].

1.3 Current Research

This paper makes a two-fold contribution to the literature. <u>First</u>, it sheds light on household bicycle ownership, a choice dimension that has hitherto been rarely studied and documented in the literature. <u>Second</u>, it involves the development of a joint model of residential neighborhood type choice and household bicycle ownership that explicitly recognizes the self-selection phenomenon described in the previous paragraph. The joint simultaneous equations model captures residential self-selection due to both observed socio demographics and unobserved attitudes and preferences. This is achieved by the use of common socio-demographic explanatory variables and common random terms in the neighborhood type choice and bicycle ownership equations. The presence of common observed variables and unobserved random terms indicates the extent to which the self-selection may be taking place. Further, the error term covariance (i.e., covariance between the common random terms) gives rise to the joint nature of neighborhood type choice and bicycle ownership. Bhat and Guo (*20*) pioneered this approach in the context of auto ownership analysis, and Pinjari et al. (*18*) used the approach in the context of commute mode choice analysis. In this paper, the joint modeling approach 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 type on bicycle ownership while accounting for residential self-selection and heterogeneity in such effects.

The paper is organized as follows. The next section describes the data. Then the model formulation is presented in Section 3. Model estimation results and policy analysis results are discussed in Sections 4 and 5, respectively. Key conclusions are presented Section 6.

2. DATA

2.1 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 sports and fitness centers, parks and gardens, restaurants, recreational businesses, , 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, a wide variety of built environment variables were extracted and/or computed for the purpose of dividing the residential neighborhoods into bicycle-friendly and less bicycle-friendly neighborhoods.

2.2 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 define a binary variable that distinguishes the TAZs of the San Francisco Bay Area into bicycle-friendly and less bicycle friendly neighborhoods. This binary variable is used as a dependent variable in the neighborhood type choice model and as an explanatory variable in the bicycle ownership model to represent bicycle friendly neighborhoods. The 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 either the clustering technique [see, for example, (*22*) and (*23*)] and or the factor analysis method [see, for example, (*24-27*)] 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 [see (*10*)]. 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.

2.3 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% of 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. (*28*)].

3. ECONOMETRIC MODELING FRAMEWORK

3.1 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.

3.1.1 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:

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

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.

3.1.2 The Bicycle Ownership Model Component

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

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

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.

3.1.3 The Joint Model System

The model system allows self-selection of households (based on their bicycle ownership preferences) into neighborhoods based on observed socio-demographics, unobserved preferences and other unobserved factors. This is achieved by the use of common socio-demographic variables and common random-error terms in the neighborhood type choice and bicycle ownership equations. The presence of common unobserved factors (captured by the common unobserved term η_q across the two equations) leads to the joint nature of the model system.

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. One can empirically test the models with both '+' and '−' signs to determine the best empirical result.

Finally, to complete the structure of Equations (1) and (2), 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 $\eta_q \sim N(0,\sigma^2)$.

3.1.4 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 common unobserved η_q term in Equations (1) and (2) is expressed as $\eta_q = \vartheta_q \exp(\iota + \varpi' v)$, where $\vartheta_q \sim N(0,1)$, ι is a constant, v is a vector of household characteristics, and ϖ is the corresponding coefficient vector. 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.

3.2 Model Estimation

Let Ω be a vector that includes all of the parameters to be estimated, c_q be a vector that vertically stacks the γ_q and δ_q vectors, and the λ_q and ϑ_q scalars, Σ be another vertically stacked vector of standard errors υ_m, ω_l, and τ. Let $\Omega_{-\Sigma}$ be 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:

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

The unconditional likelihood function can be computed for household *q* as:

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

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 [see (*29*)] are applied to approximate the multidimensional integral in Equation (5), and the resulting simulated log-likelihood function is used in the maximum likelihood estimation. Gauss matrix programming language was used to code the simulated log-likelihood functions and corresponding simulated gradients.

4. 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 neighborhood type choice and bicycle ownership
- A homogenous-joint model system of neighborhood type choice and bicycle ownership
- A disjoint (or independent) model system of neighborhood type choice and bicycle ownership
- A disjoint (or independent) model system including only a constant in the neighborhood type choice 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 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 neighborhood type 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. The negative coefficients on these three attributes appear to be unintuitive at the first glance, since one may associate households with children, living in a single-family dwelling unit, or owning a house with positive preferences for bicycle-friendly neighborhoods. This seemingly unintuitive result can be explained as follows. Recall that the neighborhood attributes may be highly correlated with each other [see (*10*)]. The factor analysis results (Section 2.2) indicate the same in the current empirical setting. The fact that six neighborhood attributes (including bicycle facility density, accessibility by bicycle mode, density measures, land-use type, and the opportunities for recreational activities by bicycle) could be collapsed into just two principal components suggests the extent of correlation between these attributes (see first block of Table 1). In other words, as one may observe from the second block of Table 1, because of the high correlation among the neighborhood attributes, bicycle-friendly neighborhoods are not only rich in bicycling facilities, connected well by the bicycle transportation network and abundant in opportunities for recreational activities involving bicycles, but also characterized by high density of street blocks and residential population. And it is possible that households with children, living in a single-family dwelling unit, or owning a house stay away from such neighborhoods with higher street block density and higher residential density. In fact, these results point to a notable finding that such households self-select to live in exclusive and sprawling sub-urban neighborhoods which also happen to be less bicycle-friendly.

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.

In view of the findings with respect to number of children, dwelling unit type, and house ownership in the context of neighborhood type choice (in the previous paragraph), it is interesting to note that such households prefer to live in less bicycle-friendly neighborhoods, but tend to own more number of bicycles. This indicates that even if such households prefer to (and thus, self-select to) live in low density neighborhoods that also happen to be less bicycle-friendly, they do prefer to own (and use) bicycles. It is possible that although the neighborhoods preferred by these households are less bicycle-friendly at a macro scale of geography (a TAZ is a neighborhood here; hence zonal level attributes define the neighborhood type), households living in such neighborhoods may create their own opportunities for bicycling within in their backyards and around their houses (i.e., in the micro scale of geography) to satisfy their bicycling preferences. Thus it is the preferences of these households formed based on such socio-demographics as presence of children, dwelling type and house ownership, that make them own (and use) more bicycles irrespective of the neighborhoods they live in. Another explanation is that 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.

The above discussion 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 a bicycle-friendly neighborhood type significantly impacts bicycle ownership in a positive way (see the last variable in the second block of Table 2). In fact, the coefficient on the neighborhood type variable ceased to be statistically insignificant when the socio-demographic variables pertaining to the number of children, dwelling type, and house ownership were dropped from the model. This is because ignoring the residential self-selection effects due to these observed attributes (i.e., ignoring their preferences to live in less bicycle-friendly neighborhoods) might have been confounded with the "true" impact of the neighborhood type. This is an important and a notable result in the context of self-selection effects. Almost all of the self-selection studies till date have reported the overestimation of neighborhood effects when the residential preferences due to socio-demographic attributes were not accounted for. In our knowledge, this is the first empirical study that indicates the possibility of an underestimation of neighborhood effects when the residential self-selection is not accounted for. Thus a general finding from this study is that ignoring residential self selection effects can potentially result in biased estimation of the neighborhood effects on travel behavior (bicycle ownership in this case). The bias could be either upward or downward depending on the specific attributes (neighborhood attributes as well as decision maker attributes) under consideration and the travel behavior context at hand.

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 due to unobserved factors (i.e., unobserved self-selection effects). It is found that heterogeneity in unobserved 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. In addition, a modest impact of income on heterogeneity in residential self-selection effects is seen.

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 unobserved residential self-selection effects. A chi-square test between the homogenous-joint model and the independent residential neighborhood type choice and bicycle ownership models with a log-likelihood ratio statistic = 0.18 with 1 degree of freedom rejects the presence of any unobserved residential self-selection in the homogenous-joint model. The heterogeneous-joint model, on the other hand, showed a statistically significant improvement in the log-likelihood. From the table, a chi-square test between the heterogeneous-joint and homogenous-joint models (a log-likelihood ratio statistic = 31.36 with 3 degrees of freedom) suggests the presence of significant variation in the unobserved residential self-selection effects in the population.

However, we would like to caution the readers that improvement in the log-likelihood should not be used as sole criterion to determine the presence of self-selection effects.

In order to further assess the presence of unobserved residential self-selection effects and corresponding heterogeneity, the t-statistic of the standard deviation of the common error component (i.e., $\sigma = \exp(\iota + \varpi' v)$) 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. Since the estimates and the t-statistics of ι and ϖ are known, the t-statistics of σ for each demographic segment (i.e., for corresponding values of v) 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 (*30*).

The t-statistics of σ for each segment were around 1.0, indicating only a marginally significant presence of residential self-selection effects in each segment. These t-statistics are, however, much 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. However, the t-statistics are not large enough to indicate a presence of significant heterogeneity in unobserved self-selection effects. The improvement in the log-likelihood and the increase in the t-statistics of σ for each segment supports the notion that one needs to account for heterogeneity in unobserved self-selection effects However, in the absence of significant unobserved residential self-selection effects to begin with (in the homogenous-joint model), 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.

5. 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.

Table 3 shows the aggregate level 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 unobserved 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). This indicates that in the current empirical context, there are no significant unobserved residential self-selection effects. Although there was a significant improvement in the log-likelihood due to capturing heterogeneity in unobserved 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 such socio-demographic variables as number of children, dwelling type, and house ownership. Thus a simple ordered logit model would have sufficed to estimate the impact of neighborhood type on bicycle ownership. 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.

6. CONCLUSIONS

Bicycle ownership and use is of much interest to the land use and transportation planning profession that is interested in promoting energy efficient and environmentally sustainable transportation mode use, and healthy lifestyles. However, there is the classic debate as to whether residential neighborhood attributes significantly impact bicycle ownership and use due to the residential self-selection effects that may be at play. It has been identified in the literature that treating built environment attributes as exogenous variables in models of travel behavior 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 neighborhood type choice and bicycle ownership that accounts for residential self-selection effects due to both observed socio-demographic attributes and unobserved factors, and accommodates unobserved heterogeneity in such effects. Using a sample of about 5000 households from the San Francisco Bay Area, a series of models are estimated using simulated maximum likelihood estimation approach. A hybrid factor- and cluster-analysis approach is used to group residential neighborhoods (denoted by traffic analysis zones) as being either bicycle-friendly or less bicycle-friendly. Model estimation results provide intuitively meaningful results.

The results indicate the presence of residential self-selection effects due to observed socio-demographic characteristics such as number of children, dwelling unit type, and house ownership. An important finding in this context is that ignoring such self-selection effects may lead to severe underestimation of the impact of bicycle-friendly neighborhood type on bicycle ownership. This is the first empirical study that reports the danger of under estimation of the neighborhood type impact on travel related behavior (bicycle ownership choice, in this paper). All other self-selection studies till date appear to have reported an overestimation of the neighborhood impacts due to ignoring residential self-selection effects.

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 unobserved residential self-selection effects or any further heterogeneity in such effects. 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 unobserved residential self-selection effects whatsoever.

In the context of the impact of bicycle-friendly neighborhoods, the policy analysis results indicate that policies aimed at modifying the built environment attributes could significantly affect bicycle ownership (and therefore, bicycle use as well). However, the effect of residential neighborhood type on bicycle ownership is found to be lower than that of socio-demographic characteristics of the household. 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 neighborhood type choice.

ACKNOWLEDGEMENTS

This research has been funded in part by Environmental Protection Agency Grant R831837. The authors acknowledge the helpful comments of five anonymous reviewers on an earlier version of the paper. The authors are also grateful to Lisa Macias for her help in typesetting and formatting this document.

REFERENCES

1. Rietveld, P. 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 2001.

2. Cao, X., S.L. Handy, and P.L. Mokhtarian. The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX. *Transportation*, Vol. 33, No. 1, 2006, pp. 1-20.

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

4. Kitamura, R., P.L. Mokhtarian, and L. Laidet. A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. *Transportation,* Vol. 24, No. 2, 1997, pp. 125-158.

5. Handy, S., X. Cao, and P.L., Mokhtarian. Self-Selection in the Relationship Between Built Environment and Walking? Evidence from Northern California. *Journal of the American Planning Association,* Vol. 72, No. 1, 2006, pp. 55-74.

6. Schwanen, T. and P.L. Mokhtarian. What If You Live in the Wrong Neighborhood? The Impact of Residential Neighborhood Type Dissonance on Distance Traveled. *Transportation Research Part D,* Vol. 10, No. 2, 2005, pp. 127-151.

7. Bagley, M. N. and P.L. Mokhtarian. The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. *Annals of Regional Science*, Vol. 36, No. 2, 2002, pp. 279-297.

8. Baltes, M.R. Factors Influencing Nondiscretionary Work Trips by Bicycle Determined from 1990 U.S. Census Metropolitan Statistical Area Data, *Transportation Research Record,* No. 1538, 1996, pp. 96-101.

9. Beck, M.J.H. and L. H. Immers. Bicycle Ownership and Use in Amsterdam. *Transportation Research Record,* No. 1441, 1994, pp. 141-148.

10. 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.

11. Ewing, R., W. Schroeer, and W. Greene. School Location and Student Travel Analysis of Factors Affecting Mode Choice, *Transportation Research Record,* No. 1895, 2005, pp. 55-63.

12. Hunt, J.D., and J.E. Abraham. Influences on Bicycle Use. *Transportation,* Vol. 34, No. 4, 2007, pp. 453-470.

13. Wigan, M.R. Bicycle Ownership, Use and Exposure: Participation and Activity Patterns in Melbourne, Australia. *Transportation Research Part A*, Vol. 18, Issues 5-6, 1984, pp. 379-398.

14. Roads and Traffic Authority (RTA). *Cycling in Sydney: Bicycle Ownership and Use.* Prepared by the Transport Data Centre, New South Wales Department of Transport, Sydney, Australia, 2007.

15. U.S. Department of Transportation and Bureau of Transportation Statistics. Bicycle and Pedestrian Data: Sources, Needs, & Gaps. BTS00-02. Washington, D.C., 2000.

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

17. Simma, A. and K.W. Axhausen. 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, 2001.

18. Pinjari, A.R., R.M. Pendyala, C.R. Bhat, and P.A. Waddell. Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice. *Transportation,* Vol. 34, No. 5, 2007, pp. 557-573.

19. Pinjari, A.R., C.R. Bhat, and D.A. Hensher. Incorporating Residential Self-Selection Effects in Activity Time-use Behavior: Formulation and Application of a Joint Mixed Multinomial Logit – Multiple Discrete-Continuous Extreme Value Model. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 2007.

20. Bhat, C.R. and J.Y. Guo. A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. *Transportation Research Part B*, Vol. 41, No. 5, 2007, pp. 506-526.

21. Cao, X., S.L. Handy, and P.L. Mokhtarian. Examining the Impacts of Residential Self-Selection on Travel Behavior: Methodologies and Empirical Findings. Presented at the 11th International Association for Travel Behavior Research (IATBR) conference, Kyoto, Japan, 2006.

22. Song, Y. and G. J. Knapp. Measuring Urban Form, *Journal of the American Planning Association,* Vol. 70, 2004, pp. 210-226.

23. Rodriguez, D.A, H.M. Young, and R.J. Schneider. 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., 2006.

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

25. Cervero, R. and K. Kockelman, Travel Demand and the 3Ds: Density, Diversity and Design. *Transportation Research Part D,* Vol. 2, No. 3, 1997, pp. 199-219.

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

27. Srinivasan, S. Quantifying Spatial Characteristics of Cities. *Urban Studies,* Vol. 39, 2002, pp. 2005-2028.

28. *Bay Area Travel Survey Final Report*, MORPACE International, Inc., Farmington Hills, MI, 2002.

29. Bhat, C. R. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research Part B*, Vol. 37, No. 9, 2003, pp. 837-855.

30. Wooldridge, J. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2002.

List of Tables

TABLE 1 Results of Factor Analysis and Cluster Analysis

Factor Analysis Results: Factor Loadings and Summary*

Factors** / **Components**	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