2013 PYKE JOHNSON AWARD: Outstanding Paper in Planning and Environment

Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects

Rajesh Paleti, Chandra R. Bhat, Ram M. Pendyala, and Konstadinos G. Goulias

Household vehicle ownership and fleet composition are choice dimensions that have important implications for policy making, particularly in the arena of energy and environmental sustainability. In the context of household vehicle ownership and type choice, it is conceivable that substantial spatial interaction effects are caused by both observed and unobserved factors. A multinomial probit model formulation is presented: it incorporates spatial spillover effects arising from both observed and unobserved factors. The model is estimated on the California add-on data set of the 2009 National Household Travel Survey. Model estimation results show that spatial dependency effects are statistically significant. The findings have important implications for model development and application in policy forecasting.

The contribution of transportation to energy consumption and greenhouse gas emissions is dependent on the nature of vehicular travel undertaken by households. The number of vehicles owned, the types of vehicles owned (size, weight, fuel type, and age), and the extent to which different vehicles are used (miles of travel) are all key determinants of energy consumption and greenhouse gas emissions. In the past 25 years, the split between cars and light-duty trucks in the nation's vehicle fleet has changed dramatically; whereas light-duty trucks (including pickup trucks, minivans, and sport utility vehicles) accounted for just about 20% of the fleet 25 years ago, they now account for about half of all vehicles on the nation's roadways (*1*). This dramatic shift in vehicular fleet composition and utilization has had far-reaching energy and environmental consequences.

The impact of the composition and utilization of the household vehicular fleet on energy consumption and greenhouse gas emissions calls for the incorporation of behavioral models of vehicle type choice and utilization in transportation demand forecasting models. Such models would provide the ability to forecast energy and environmental impacts of shifting vehicle ownership and utilization patterns arising from alternative policy decisions, the advent of alternative fuel vehicle technologies, and changes in household and personal vehicular preferences. In this context, although there are several earlier efforts on vehicle ownership analysis in the literature, much remains to be done in development of behavioral models of household vehicle fleet composition and utilization choices and in connecting such choices to energy and emissions estimates.

A particular issue that has not been adequately addressed in the literature on vehicle ownership and utilization is that there may be spatial interaction effects in household vehicle ownership and type choice that are both observed and unobserved. Vehicle choices that households make are likely to be influenced by interactions with neighboring households and the choices that neighboring households make. If a household observes that many of its neighbors own and drive hybrid electric vehicles, or hears good reviews about such vehicles from neighbors who already own and drive them, then the household may be motivated and influenced to also own and drive a hybrid electric vehicle. Spatial interaction effects may arise from unobserved attitudinal preferences in that households with similar lifestyle preferences cluster in neighborhoods with attributes of the built environment that are conducive to their lifestyle choices.

This paper contributes to vehicle ownership and fleet composition analysis by presenting a multinomial probit model that explicitly accounts for spatial interaction effects in these choice phenomena. Underlying the multinomial probit model with spatial interaction effects is a behavioral framework that estimates not only the number of vehicles owned by a household but also the vehicle type choice, allowing construction of the entire vehicle fleet for a household while spatial dependency effects are explicitly considered.

SPATIAL DEPENDENCE IN CHOICE MODELING

There has been increasing attention in the past decade on accommodating spatial dependency effects in modeling of choice-making behaviors by agents in a variety of contexts (*2*). Several efforts have applied spatial correlation structures developed for modeling continuous dependent variables in the context of discrete choice models of behavior (*2, 3*). However, these efforts have been hampered by the need for evaluation of multidimensional integrals of the order of the product of the number of decision agents and the number of alternatives minus one for unordered multinomial response choice models.

Several studies have sidestepped the high-dimensional problem inherent in global and general spatial dependency structures by

R. Paleti and C. R. Bhat, Department of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, 301 East Dean Keeton Street, Stop C1761, Austin, TX 78712-1172. R. M. Pendyala, School of Sustainable Engineering and the Built Environment, Arizona State University, Room ECG252, Tempe, AZ 85287-5306. K. G. Goulias, Department of Geography, University of California, Santa Barbara, CA 93106-4060. Alternate affiliation for C. R. Bhat: King Abdulaziz University, Jeddah 21589, Saudi Arabia. Corresponding author: C. R. Bhat, bhat@mail.utexas.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2343, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 86–94.
DOI: 10.3141/2343-11

assuming that the dependency originates only from observed exogenous covariates of proximate decision agents (*4, 5*). However, this is untenable in the context of several choice situations in which spatial dependence naturally arises from didactic interactions between decision agents. Households may be viewed as developing utilities (or preferences) for vehicle choice alternatives on the basis of a set of observed factors (such as income and presence of children in neighboring households) as well as unobserved tastes, attitudes, and location factors (such as how green a household is in its views and whether there are continuous sidewalks or bicycle paths in the neighborhood). The utility vector of one household is likely to be influenced by the utility vector of nearby households because of didactic interactions and interchanges (in which utility signals are bounced across decision agents). In this process, there is a spatial spillage effect not only based on the observed covariate effects of neighboring households but also caused by unobserved factors. For example, a neighboring household's perception of greenness or the quality of sidewalks and bike paths may spill over and influence choices of another household. Further, residential self-selection effects may lead to a sorting of households according to similarity in unobserved preferences in vehicle type choice.

In discrete choice models, ignoring these spillage effects caused by observed factors or unobserved factors will lead to inconsistent estimates of the effects of observed covariates. As indicated by Anselin, it behooves the analyst to include spatial spillover effects in both the observed covariates and the errors unless there are strong a priori reasons not to do so (*6*). In the present paper, a spatial lag formulation is adopted to accommodate global spatial dependence effects (caused by both observed covariate and error spillage effects) in household decisions on vehicle type choice. The specific model structure and formulation implemented in this paper allow modeling of the entire vehicle fleet composition of households, discussed in the next section. Development of a multinomial probit model with continuous spatial dependency effects (caused by both observed and unobserved factors) that can model the entire vehicle fleet composition is the novel contribution of this paper.

DATA

The data set used in this study was derived from the California add-on component of the 2009 National Household Travel Survey. The National Household Travel Survey is conducted by the U.S. Department of Transportation to measure the amount of personal travel that is undertaken by the populace. Individual states and metropolitan areas may purchase and commission additional data collection within their jurisdictions if they want larger samples for their own analysis and planning applications. A subsample from the Los Angeles region within the California add-on survey sample was extracted for the analysis conducted in this paper. Because spatial interaction effects are likely to be more localized, a data set from a smaller geographic region was used. The choice to limit the sample size (and thus avoid inflated *t*-statistics that might arise from the use of large samples) was another consideration in the selection of a subsample. Finally, the selection of this specific subsample made it possible to merge accessibility measures and land use data at the census tract level that have been compiled in connection with an ongoing parallel effort to develop a comprehensive activity-based microsimulation model system for the Southern California Association of Governments (*7*). The accessibility measures are opportunity-based indicators that measure the number of activity opportunities by 12 industry types and by total roadway length of roadway types that can be reached within 10 min by automobile from the home census tract during the morning peak period (6 to 9 a.m.).

The data set includes detailed socioeconomic and demographic data for individuals and households, along with information about the vehicle fleet in each household. After extensive cleaning and filtering for missing data, a survey sample of 961 households was available for analysis. To limit the sample size and for computational tractability, a 25% random sample of 243 households residing in 200 census tracts was chosen for model estimation. For the model estimation exercise in this paper, vehicle type choice was represented as a combination of two dimensions, body type and vintage. Two body types were considered: car and noncar (encompassing sport utility vehicles, vans and minivans, and pickup trucks). Two age categories were considered: less than or equal to 5 years old and more than 5 years old. Thus four vehicle type alternatives are defined in this paper.

The descriptive characteristics of the sample of 243 households indicate that the data set is suitable for the model estimation. It was found that 8.2% of households have no vehicle, another 34.5% have one vehicle, and 40% have two vehicles. Among the vehicles in the sample, 40% are old cars, 24% are new cars (less than or equal to 5 years old), another 24% are old noncars, and 12% are new noncars. For other descriptive statistics, 82% of the households are of non-Hispanic origin and 68% of individuals report their race as Caucasian. About 70% of households own the home in which they reside. It was found that one-fifth of the households report an annual income less than \$20,000, and an equal proportion report incomes of between \$20,000 and \$45,000. About 38% of the households report annual income of greater than \$75,000. About 47% of the households report having one adult, and another 46% report having two adults. Nearly 34% of households have zero workers, and 44% have one worker. About 17% of households report having one self-employed individual. There is one person with more than one job in 11% of the households. The employed individuals report a mean distance to work of 6.1 mi. Only 1% of the households report having a child up to 5 years of age, but 12% of households report having a child 6 to 10 years of age. About 12% of households report having a child 11 to 15 years of age (households are not necessarily mutually exclusive). Just over one-third of households report having an adult who is 65 years of age or older. About 35% of households are immigrant households. The mean distance between households (based on the census tracts of household residences), which is the distance measure used to capture spatial dependence effects caused by proximity, is 11.1 mi with a standard deviation of 6.6 mi. The corresponding median distance is 11 mi. Additionally, 20.4% of household pairings have interhousehold distances of less than 5 mi. Thus, there are enough households close to one another, as well as enough variation in the interhousehold distances across household pairings, for estimation of spatial dependency effects.

MODELING METHODOLOGY

The behavioral framework adopted in this study assumes that the observed vehicle fleet of a household is the result of a series of unobserved (to the analyst) repeated synthetic discrete choice occasions in which the household chooses not to purchase a vehicle or chooses a vehicle of a certain type. The number of synthetic choice occasions in such a vertical (over time) choice setting is linked to the number of driving-age members in the household to exploit that the number of vehicles owned by a household is virtually never greater than the number of driving age members (say, N) plus two. (In the data set used in the present analysis, 99.1% of households were covered by this condition.) Thus, for each household, a set of $N + 2$ synthetic choice occasions is created and an appropriate choice is assigned as the dependent variable. For estimation, a procedure is needed for assigning a chosen alternative at each synthetic occasion. For this, the temporal sequence of vehicle purchases of the household, as reported in the survey, is used. For example, a household owns an old sedan and a new sport utility vehicle, and the old sedan was purchased first. The old sedan is the chosen alternative at the first choice occasion, and the new sport utility vehicle is the chosen alternative in the second. The chosen alternative in the remaining two choice occasions is "no vehicle purchased." For the second choice occasion, information that the household already has an old sedan is used as an explanatory variable. It is also possible to assign the old sedan to the first choice occasion, no vehicle in the second, no vehicle in the third, and the new sport utility vehicle in the fourth occasion. However, both assignments give the same results, because the dynamics are based on what the household already owns in total, not what was chosen in the immediate previous choice occasion.

The described procedure mimics the dynamics of fleet ownership decisions, although there is no temporal component of the dynamics involved because only synthetic choice occasions are considered; the available observed information is only that of vehicles held at a cross-sectional point in time with information on the sequence in which the currently held vehicles were purchased. The approach here is not a true model of vehicle fleet evolution that analyzes the dynamics of vehicle transaction decisions over time. The estimation of such evolution models, although appealing from a behavioral perspective, has been hampered by the lack of longitudinal data on vehicle transactions. Moreover, many dynamic models have focused on vehicle ownership (i.e., transactions) with inadequate emphasis on vehicle type, usage, and vintage considerations of the household fleet.

Modeling Approach

Let the instantaneous utility U_{qti} of household q ($q = 1, 2, \ldots, Q$) at synthetic choice occasion t ($t = 1, 2, \ldots, T_q$) for vehicle type choice i ($i = 1, 2, \ldots, I$; $I = 5$ in the empirical context of the present paper, including the "no vehicle purchase" alternative) be a function of a ($K \times 1$)-column vector of exogenous attributes $\mathbf{x}_{qti}$ (including household demographics, types of vehicles chosen before the tth choice occasion, and activity-travel environment characteristics). Let $T_q = N_q + 2$, where T_q is the number of synthetic choice occasions for household q and N_q is the number of driving-age members in household q. t does not have a chronological time interpretation. It is simply a device for accommodating multiple synthetic choice occasions and mimics the dynamics of fleet ownership decisions. That is, $t = 1$ for household A does not have any chronological time bearing to $t = 1$ or $t = 2$ for neighboring household B. However, the choice occasions of different households may be considered to occur during a period in which households are interacting and exchanging utility signals. Thus, the spatial dependence across households is specified for each vehicle type i without any specific association to the choice occasion. That is, the utility U_{qti} for household q at choice occasion t for alternative i is related to the utility $U_{q't'i}$ for household q' and alternative i at each (and all) of the choice occasions t' ($t' = 1, 2, \ldots, T_{q'}$) of household q'. This is an important distinction from the traditional spatial dependency specifications for spatial panel discrete choice models and leads to a specific form for the model in this study that has not appeared previously in the literature.

The utility U_{qti} incorporating a spatial lag structure is written as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} \sum_{t'=1}^{T_{q'}} U_{q't'i} + \tilde{\alpha}_{qi} + \boldsymbol{\beta}_q' \mathbf{x}_{qti} + \tilde{\varepsilon}_{qti} \quad (1)$$

where

$w_{qq'}$ = distance-based spatial weight corresponding to units q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$) for each (and all) q,
δ ($0 < \delta < 1$) = spatial lag autoregressive parameter,
$\tilde{\alpha}_{qi}$ = normal random-effect term capturing household-specific stationary preference effect for vehicle type i, and
$\boldsymbol{\beta}_q$ = household-specific ($K \times 1$)-vector of coefficients assumed to be a realization from multivariate normal distribution with mean vector $\mathbf{b}$ and covariance $\tilde{\boldsymbol{\Omega}} = \mathbf{LL}'$.

Let $\boldsymbol{\beta}_q = \mathbf{b} + \breve{\boldsymbol{\beta}}_q$, where $\breve{\boldsymbol{\beta}}_q \sim \text{MVN}_K(0, \tilde{\boldsymbol{\Omega}})$ is a mixing (multivariate) distribution to capture unobserved sensitivity variations (to the exogenous variables in the vector $\mathbf{x}_{qti}$) across households. (MVN_K represents the multivariate normal distribution of dimension K.) Write $\tilde{\alpha}_{qi} = \tilde{\alpha}_i + \breve{\alpha}_{qi}$, and let the mean and variance–covariance matrix of the vertically stacked ($I \times 1$)-vector of random-effect terms $\tilde{\boldsymbol{\alpha}}_q$ [$= (\tilde{\alpha}_{q1}, \tilde{\alpha}_{q2}, \ldots, \tilde{\alpha}_{qI})'$] be $\tilde{\mathbf{A}}$ and $\tilde{\boldsymbol{\Lambda}}$, respectively. $\tilde{\varepsilon}_{qti}$ in Equation 1 is a normal error term uncorrelated with $\breve{\boldsymbol{\beta}}_q$ and all $\tilde{\alpha}_{qi}$ terms ($i = 1, 2, \ldots, I$) and also uncorrelated across observation units q and synthetic choice occasions t. However, at each synthetic choice occasion t for household q, the $\tilde{\varepsilon}_{qti}$ terms may have a covariance (dependency) structure across vehicle types i because of unobserved factors for choice occasion that simultaneously increase or simultaneously decrease the utility of certain types of vehicles:

$$\tilde{\boldsymbol{\varepsilon}}_{qt}\left[= \left(\tilde{\varepsilon}_{qt1}, \tilde{\varepsilon}_{qt2}, \ldots, \tilde{\varepsilon}_{qtI}\right)'\right] \sim \text{MVN}_I\left(0, \tilde{\boldsymbol{\Psi}}\right)$$

As usual, appropriate scale and level normalization must be imposed on $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$ for identification purposes. Specifically, only utility differentials matter in discrete choice models. At the same time, whenever utility differentials are taken during estimation, they must all originate from the same underlying matrices $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$. To achieve this, take the utility differentials with respect to the first alternative. Then, only the elements $\alpha_{qi1} = \tilde{\alpha}_{qi} - \tilde{\alpha}_{q1}$ ($i \neq 1$) and its covariance matrix $\boldsymbol{\Lambda}_1$, and the covariance matrix $\boldsymbol{\Psi}_1$ of $\xi_{qti1} = \tilde{\varepsilon}_{qti} - \tilde{\varepsilon}_{qt1}$ ($i \neq 1$), are estimable. A normalization $\tilde{\alpha}_{q1} = 0 \; \forall \; q$ is applied, implying that $\tilde{a}_1 = 0$. Also, develop $\boldsymbol{\Lambda}$ from $\boldsymbol{\Lambda}_1$ by adding an additional row at the top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. Similarly, construct $\boldsymbol{\Psi}$ from $\boldsymbol{\Psi}_1$ by adding a row at the top and a column to the left. This first row and the first column of the matrix $\tilde{\boldsymbol{\Psi}}$ are also filled with zero values. An additional normalization must be imposed on $\tilde{\boldsymbol{\Psi}}$ because the scale is also not identified. For this, normalize the element of $\tilde{\boldsymbol{\Psi}}$ in the second row and second column to the value of one. All these normalizations are needed for econometric identification.

Next, define the following:

$$\mathbf{U}_{qt} = \left(U_{qt1}, U_{qt2}, \ldots, U_{qtI}\right)', \tilde{\varepsilon}_{qt} = \left(\tilde{\varepsilon}_{1t1}, \tilde{\varepsilon}_{qt2}, \ldots, \tilde{\varepsilon}_{qtI}\right)' \; (I \times 1 \text{ vectors})$$

$$\mathbf{U}_q = \left(\mathbf{U}'_{q1}, \mathbf{U}'_{q2}, \ldots, \mathbf{U}'_{qT_q}\right)', \tilde{\boldsymbol{\varepsilon}}_q = \left(\tilde{\boldsymbol{\varepsilon}}'_{q1}, \tilde{\boldsymbol{\varepsilon}}'_{q2}, \ldots, \tilde{\boldsymbol{\varepsilon}}'_{qT_q}\right)' \left((T_q \times I) \times 1 \text{ vectors}\right)$$

$$\mathbf{U} = \left(\mathbf{U}'_1, \mathbf{U}'_2, \ldots, \mathbf{U}'_Q\right)', \tilde{\boldsymbol{\varepsilon}} = \left(\tilde{\boldsymbol{\varepsilon}}'_1, \tilde{\boldsymbol{\varepsilon}}'_2, \ldots, \tilde{\boldsymbol{\varepsilon}}'_Q\right)' (RI \times 1 \text{ vectors}), R = \sum_{q=1}^{Q} T_q$$

$$\breve{\boldsymbol{\alpha}}_q = \left(\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, \ldots, \breve{\alpha}_{qI}\right)' (I \times 1 \text{ vector}),$$

$$\breve{\boldsymbol{\alpha}} = \left[\left(1_{T_1} \otimes \breve{\alpha}_1\right)', \left(1_{T_2} \otimes \breve{\alpha}_2\right)', \ldots, \left(1_{T_Q} \otimes \breve{\alpha}_Q\right)'\right]' (RI \times 1 \text{ vector})$$

$$\mathbf{x}_{qt} = \left(\mathbf{x}_{qt1}, \mathbf{x}_{qt2}, \ldots, \mathbf{x}_{qtI}\right)' (I \times K \text{ matrix}), \mathbf{x}_q = \left(\mathbf{x}'_{q1}, \mathbf{x}'_{q2}, \ldots, \mathbf{x}'_{qT_q}\right)' \left((T_q \times I) \times K \text{ matrix}\right)$$

$$\mathbf{x} = \left(\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_Q\right)' (RI \times K \text{ matrix})$$

$$\breve{\boldsymbol{\beta}} = \left(\breve{\boldsymbol{\beta}}'_1, \breve{\boldsymbol{\beta}}'_2, \ldots, \breve{\boldsymbol{\beta}}'_Q\right)' (QK \times 1 \text{ vector})$$

Let $\mathbf{IDEN}_E$ be the identity matrix of size E, let $\mathbf{1}_E$ be a column vector of size E with all its elements taking the value of one, and let $\mathbf{1}_{EE}$ be a square matrix of size E with all unit elements. Also, define the following matrix:

$$\tilde{\mathbf{x}} = \begin{bmatrix} x_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & x_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & x_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & x_Q \end{bmatrix} (RI \times QK \text{ matrix}) \quad (2)$$

Let $\tilde{R}_q = \Sigma_{j=1}^{q-1} T_j$, with the convention that $\tilde{R}_1 = 0$, and let $\breve{R}_q = \tilde{R}_q \times I$. Define a matrix $\mathbf{C}$ of size $RI \times RI$ that is filled with submatrices of size $(T_q \times I) \times (T_{q'} \times I)$ as follows:

$$[\mathbf{C}]_{\{(\breve{R}_q+1)-\breve{R}_{q+1}\}\{(\breve{R}_{q'}+1)-\breve{R}_{q'+1}\}} = w_{qq'} \otimes \mathbf{1}_{T_q T_{q'}} \otimes \mathbf{IDEN}_I$$

where

$\mathbf{S} = [\mathbf{IDEN}_{RI} - \delta\mathbf{C}]^{-1}$ ($RI \times RI$ matrix)

$\breve{\mathbf{A}} = [(1_{T_1} \otimes \tilde{\mathbf{A}})', (1_{T_2} \otimes \tilde{\mathbf{A}})', \ldots, (1_{T_Q} \otimes \tilde{\mathbf{A}})']'$ ($RI \times 1$ matrix)

$[\mathbf{C}]_{\{(\breve{R}_q+1)-\breve{R}_{q+1}\},\{(\breve{R}_{q'}+1)-\breve{R}_{q'+1}\}} =$ submatrix of $\mathbf{C}$ that corresponds to $(\breve{R}_q + 1)$th through $\breve{R}_{q+1}$th rows and $(\breve{R}_{q'} + 1)$th through $\breve{R}_{q'+1}$th columns.

Then, Equation 1 may be written in matrix notation as

$$\mathbf{U} = \mathbf{S}\left[\breve{\mathbf{A}} + \mathbf{x}\mathbf{b} + \breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\right] \quad (3)$$

Let $[\cdot]_e$ indicate the eth element of the column vector $[\cdot]$, and let $d_{qti} = \breve{R}_q + (t-1)I + i$. Equation 3 can be equivalently written as

$$U_{qti} = \left[\mathbf{S}\left\{\breve{\mathbf{A}} + \mathbf{x}\mathbf{b}\right\}\right]_{d_{qti}} + \left[\mathbf{S}\left\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\right\}\right]_{d_{qti}} \quad (4)$$

Define

$$V_{qti} = \left[\mathbf{S}\left\{\breve{\mathbf{A}} + \mathbf{x}\mathbf{b}\right\}\right]_{d_{qti}}$$

and

$$\varepsilon_{qti} = \left[\mathbf{S}\left\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\right\}\right]_{d_{qti}}$$

Household q chooses the vehicle type at synthetic choice occasion t that provides maximum utility. Let the chosen vehicle type (assigned as described previously) for household q at occasion t be m_{qt}. In the utility differential form, Equation 4 may be written as

$$y_{qtim_{qt}} = U_{qti} - U_{qtm_{qt}} = H_{qtim_{qt}} + \xi_{qtim_{qt}}, H_{qtim_{qt}} = V_{qti} - V_{qtm_{qt}}, \text{ and } \xi_{qtim_{qt}} = \varepsilon_{qti} - \varepsilon_{qtm_{qt}}, i \neq m_{qt} \quad (5)$$

Then stack the utility differentials $y_{qtim_{qt}}$ ($= U_{qti} - U_{qtm_{qt}}$, $i \neq m_{qt}$) in the following order: $\mathbf{y}_{qt} = (y_{qt1m_{qt}}, y_{qt2m_{qt}}, \ldots, y_{qtIm_{qt}})'$, an $(I-1) \times 1$ vector; $\mathbf{y}_q = (\mathbf{y}'_{q1}, \mathbf{y}'_{q2}, \ldots, \mathbf{y}'_{qT_q})'$, an $[(I-1) \times T_q] \times 1$ vector; and $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_Q)'$, an $[(I-1) \times R] \times 1$ vector. Correspondingly, let $\mathbf{H}_{qt} = (H_{qt1m_{qt}}, H_{qt2m_{qt}}, \ldots, H_{qtIm_{qt}})'$, an $(I-1) \times 1$ vector; $\mathbf{H}_q = (\mathbf{H}'_{q1}, \mathbf{H}'_{q2}, \ldots, \mathbf{H}'_{qT_q})'$, an $[(I-1) \times T_q] \times 1$ vector; and $\mathbf{H} = (\mathbf{H}'_1, \mathbf{H}'_2, \ldots, \mathbf{H}'_Q)'$, an $[(I-1) \times R] \times 1$ vector. $\mathbf{y}$ has a mean vector $\mathbf{H}$.

To determine the covariance matrix of $\mathbf{y}$, several additional matrix definitions are needed. Define a matrix $\boldsymbol{\Lambda}$ of size $RI \times RI$ that is block diagonal with each block diagonal as follows: $[\boldsymbol{\Lambda}]_{\{(\breve{R}_q+1)-\breve{R}_{q+1}\},\{(\breve{R}_{q+1})-\breve{R}_{q+1}\}} = \mathbf{1}_{T_qT_q} \otimes \tilde{\boldsymbol{\Lambda}}$ $(q = 1, 2, \ldots, Q)$, $\boldsymbol{\Omega} = \tilde{\mathbf{x}}(\mathbf{I}_Q \otimes \tilde{\boldsymbol{\Omega}})\tilde{\mathbf{x}}'$ ($RI \times RI$ matrix), and $\boldsymbol{\Psi} = \mathbf{IDEN}_R \otimes \tilde{\boldsymbol{\Psi}}$ ($RI \times RI$ matrix). Let $\tilde{\mathbf{F}} = \mathbf{S}[\boldsymbol{\Lambda} + \boldsymbol{\Omega} + \boldsymbol{\Psi}]\mathbf{S}'$ and define $\mathbf{M}$ as an $(q = 1, 2, \ldots, Q)$: $[(I-1) \times R] \times [I \times R]$ block diagonal matrix with each block diagonal having $(I-1)$ rows and I columns corresponding to the tth synthetic choice occasion of household q. This $(I-1) \times I$ matrix for household q and choice occasion t corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the m_{qt}th column. Finally, the multivariate distribution of the utility differentials is obtained, $\mathbf{y} : \mathbf{y} \sim \mathbf{MVN}\ (\mathbf{H}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{M}\tilde{\mathbf{F}}\mathbf{M}'$. Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\mathbf{b}'$; Vech$(\tilde{\boldsymbol{\Omega}})$; $\tilde{\mathbf{A}}'$, Vech$(\tilde{\boldsymbol{\Lambda}})$,Vech$(\tilde{\boldsymbol{\Psi}})$, $\delta]'$, where Vech$(\tilde{\boldsymbol{\Omega}})$ represents the row vector of upper triangle elements of $\tilde{\boldsymbol{\Omega}}$. Then, the likelihood of the observed sample may be written succinctly as Prob$[\mathbf{y}^* < \mathbf{0}]$.

$$L_{ML}(\boldsymbol{\theta}) = \text{Prob}\left[\mathbf{y}^* < 0\right] = F_{R\times(I-1)}\left(-\mathbf{H}, \boldsymbol{\Sigma}\right) \quad (6)$$

where $F_{R\times(I-1)}$ is the multivariate cumulative normal distribution of $R \times (I-1)$ dimensions.

Despite advances in simulation techniques and computational power, the evaluation of such a high-dimensional integral is infeasible with established estimation techniques.

Model Estimation Procedure

In view of the computational intractability of the likelihood function presented earlier, the present study uses Bhat's maximum approximate composite marginal likelihood (MACML) inference approach in the estimation (*8*). The MACML approach combines a CML estimation approach with an approximation method for evaluating the multivariate standard normal cumulative distribution function. The CML approach works as follows. Instead of developing the likelihood of the entire sample, consider developing a surrogate likelihood function that is the product of the probability of easily computed marginal events. For instance, one may compound (multiply) pairwise probabilities of household q choosing the actual chosen vehicle type m_{qt} at occasion t and choosing the actual chosen vehicle type m_{qs} at occasion s, of household q choosing vehicle type m_{qt} at occasion t and household q' choosing vehicle type $m_{q's}$ at time s, and so on. The CML estimator is then the one that maximizes the compounded probability of all pairwise events. The CML function may be written as

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q} \prod_{q'=q}^{Q} \prod_{t=1}^{T} \prod_{t'=t}^{T} \text{Prob}\left(C_{qt} = m_{qt}, C_{q't'} = m_{q't'}\right) \quad \text{with } q \neq q' \text{ when } t = t' \qquad (7)$$

where C_{qt} is an index for the vehicle type chosen by household q at occasion t. Each of these pairwise probabilities is of $(I-1) \times 2$ dimensions, which may be computed easily with the multivariate standard normal cumulative distribution approximation method embedded in the MACML method. The pairwise marginal likelihood function of Equation 7 comprises $R(R-1)/2$ pairs of multivariate pairwise probability computations, which can become quite time-consuming. However, in a spatial-temporal case where spatial dependency drops quickly with interobservation distance, it should suffice to retain pairs within a certain threshold distance. This threshold value is estimated by testing different distance bands, starting from a small distance band and increasing the band. Then, the asymptotic variance matrix $V_{\text{CML}}(\hat{\boldsymbol{\theta}})$ is estimated for each distance band, and the threshold distance value (say, $\tilde{d}_{\text{thresh}}$) is chosen as the value beyond which there is either an increase or no additional decrease in the total variance across all parameters as given by $tr[V_{\text{CML}}(\hat{\boldsymbol{\theta}})]$ (i.e., the trace of the matrix $[V_{\text{CML}}(\hat{\boldsymbol{\theta}})]$.

The CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of Godambe's sandwich information matrix (*9*). Bhat provided details about how to compute the covariance matrix (*8*).

In the estimations, the positive-definiteness of each of the $\tilde{\boldsymbol{\Omega}}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$ matrices is guaranteed by writing the logarithm of the pairwise likelihood in terms of the Cholesky-decomposed elements of these matrices and maximizing with respect to these elements of the Cholesky factor. To ensure the constraint $0 < \delta < 1$, this term is parameterized as $\delta = 1/[1 + \exp(\tilde{\delta})]$. Once estimated, the $\tilde{\delta}$ estimate can be translated back to obtain estimates of δ.

ESTIMATION RESULTS

This section presents results of the estimation of the multinomial probit model with spatial dependency effects on the California add-on data set of the 2009 National Household Travel Survey. Estimation results are presented in Table 1. Several model specifications were estimated before the final model specification was arrived at. None of the mixing parameters proved to be statistically significant in the final model specification. This result indicates that there is no significant household-specific heterogeneity in the variable effects on the vehicle type decisions. Even in the absence of mixing on variables, the model does not collapse to a cross-sectional spatial model. This is because the setup of the model is such that the utility associated with an alternative for one household at any given synthetic choice occasion is influenced by the utility associated with the same alternative across all synthetic choice occasions of all other households in the region, which leads to a pseudo unbalanced panel setup because of an unequal number of choice occasions across individuals. Another key finding is that there were no significant deviations in the error covariance matrix $\boldsymbol{\Psi}_1$ of $\xi_{qti1} = \tilde{\varepsilon}_{qti} - \tilde{\varepsilon}_{qt1}$ $(i \neq 1)$ from the corresponding matrix in an independent multinomial probit (MNP) model. The covariance matrix of the error term differences in an independent MNP model has 1 and 0.5 as the diagonal and off-diagonal elements, respectively. This finding implies that at any given synthetic choice occasion the occasion-specific unobserved factors that influence the utility associated with vehicle type alternatives are all independent and identically distributed.

Although the magnitudes of the constants cannot be directly compared across alternatives (because there are continuous variables in the utility formulations), the relative values may be loosely interpreted as indicative of the baseline preference for vehicle types. It appears that new cars are the least preferred vehicle type, and old cars are the most preferred; there is little difference in the baseline preference between old and new noncars. Households with higher levels of education are less inclined to acquire new noncars. It is possible that these households are more environmentally conscious, and savvy consumers shun the expense and environmental consequences of driving new noncars (sport utility vehicles, trucks, and vans). Those of Hispanic origin show a greater inclination to acquire old noncars, and African Americans are less likely to acquire old cars and new noncars.

Households that own their homes have a greater utility for all vehicle alternatives (in comparison with those that do not own their homes) and have a particular preference for new vehicles. As expected, households in the highest income category have a positive utility for all vehicle alternatives, with a higher preference for new noncars and the lowest preference for old noncars. However, the difference in magnitudes of coefficients across vehicle type alternatives is modest. As the number of adults increases, households are more likely to acquire old cars, new cars, or old noncars—it is assumed that these households have a greater need for multiple cars and show a greater disinclination to acquire new noncars because of budget constraints. However, with the easing of budget constraints that comes with a greater number of workers in the household, households show a greater preference for new cars or noncars and shun older cars.

As the mean distance to work increases, households are more likely to acquire new noncars, likely because people are looking for larger vehicles that are more comfortable and reliable for the longer commute. The presence of children is associated with a smaller likelihood of acquiring older cars and a greater likelihood of acquiring older noncars. It is assumed that such households prefer larger vehicles for space and newer cars for reliability. Immigrant households have a lower utility across all vehicle type alternatives compared with nonimmigrant households but have a smaller negative coefficient on the noncar alternatives. Immigrant households may be located in more dense neighborhoods and may be more walking and transit oriented, which would contribute to the lower utility across all vehicle types. The coefficients on the noncar alternatives are less negative, probably because these households are of larger size, which motivates the acquisition of noncars in preference to cars. Households with senior adults are less likely to acquire noncars. There are two possible explanations for this: seniors may have diminishing skills that make driving and controlling larger vehicles cumbersome, and seniors may be living in smaller households (empty nests) and so do not need larger vehicles (noncars).

Because the temporal sequence in which a household acquired vehicles is known in the survey data set, information about the existing vehicle fleet was used as explanatory variables in the utility specification of vehicle type alternatives for all choice occasions after the first. This specification mimics the underlying dynamics in purchase decisions when existing vehicles in the household influence the vehicles that households acquire subsequently. As expected, parameter estimates for all vehicle types are negative, suggesting that households increasingly choose to acquire no vehicle as they build their fleets. The relative magnitudes of the coefficients can be used to draw inferences about how households tend to construct their fleets. It appears that households are somewhat variety seeking. For example, as the number of old cars increases, households are more likely to

TABLE 1 Spatial Model Estimation Results for Vehicle Type Choice

	Old Car		New Car		Old Noncar		New Noncar	
Variable	Coeff.	*t*-Stat.	Coeff.	*t*-Stat.	Coeff.	*t*-Stat.	Coeff.	*t*-Stat.
Constant	−0.0351	−0.19	−0.6960	−3.64	−0.5647	−1.92	−0.5432	−1.94
Demographics								
Highest education attainment in household (base is college degree or less)								
Bachelor's degree	—	—	—	—	—	—	−0.5185	−9.12
Graduate	—	—	—	—	—	—	−0.4172	−6.48
Hispanic status (base category is non-Hispanic): Hispanic origin	—	—	—	—	0.4983	5.81	—	—
Race (base category is all other races): African American	−0.3183	−6.35	—	—	—	—	−0.1993	−4.22
Housing tenure (base category is rental home): own	0.3690	9.48	0.6321	11.08	0.7265	11.71	1.0050	17.81
Household income (base category is all lower-income levels): greater than $75,000	0.7863	12.91	0.6581	8.89	0.5848	8.29	0.8284	11.80
Number of adults	0.5195	15.57	0.3884	8.27	0.6625	14.02		
Number of full-time workers	—	—	0.4139	8.24	—	—	0.4485	9.00
Number of people with more than one job	—	—	−0.3947	−4.12	—	—		
Mean distance to work (mi)	—	—	—	—	—	—	0.0115	9.97
Presence of children								
6–10 years	−0.2136	−4.48	—	—	—	—	—	—
11–15 years	—	—	—	—	0.1985	2.07	—	—
Presence of senior adults	—	—	—	—	−0.5534	−7.87	−0.5617	−8.18
Presence of individual with prolonged medical condition (>5 years)	—	—	−0.3844	−9.21	—	—	—	—
Immigration status (base is nonimmigrant household): immigrant household	−0.3733	−6.39	−0.3733	−6.39	−0.1957	−3.62	−0.1957	−3.62
Existing Vehicle Fleet Characteristics								
Number of old cars	−1.1195	−12.84	−0.9968	−18.36	−1.0363	−14.35	−0.8437	−12.52
Number of new cars	−2.0845	−32.36	−1.2579	−16.29	−2.3589	−9.24	−0.7159	−10.11
Number of old noncars	−0.9627	−9.96	−0.7004	−11.51	−1.3573	−15.60	−0.4919	−10.52
Number of new noncars	−1.5708	−34.02	−1.1944	−8.93	−1.9237	−30.98	−1.7878	−22.35
Accessibility Measures								
Primary arterial roads roadway length within 10 min (mi/10^4)	—	—	—	—	−1.7816	−12.15	—	—
Minor arterial roads roadway length within 10 min (mi/10^4)	—	—	—	—	−0.9369	−6.03	—	—
Collector roads roadway length within 10 min (mi/10^4)	—	—	—	—	0.5306	16.06	—	—
Total manufacturing employment that can be reached within 10 min (/10^4)	—	—	−2.4896	−2.20	—	—	—	—
Total arts employment that can be reached within 10 min (/10^4)	−9.2334	−9.33	—	—	—	—	—	—

NOTE: Spatial interaction parameter (δ) = 0.1872; *t*-statistic = 3.80. — = variable has no significant impact on the alternative; coeff. = coefficient; *t*-stat. = *t*-statistic.

acquire new noncars (least negative coefficient); as the number of new cars increases, households are more likely to acquire new noncars (and shun older cars); as the number of old noncars increases, households are more likely to acquire new noncars or new cars; and finally, as the number of new noncars increases, households are more likely to acquire new cars. In all cases, the least negative coefficient is associated with a car type different from that representing the explanatory variable for the existing vehicle fleet.

Households with good access to primary and minor arterials have a lower preference for older noncars, which suggests that these households may be more auto oriented (and hence located in census tracts with good roadway presence) and prefer to drive newer cars. Households in census tracts closer to manufacturing employment are less likely to acquire new cars, possibly because these census tracts are in lower-income areas and new car purchases are thus challenging. Households in census tracts close to arts employment are less likely to acquire old cars; it is possible that these census tracts are in trendy urban arts districts and people who locate there are more likely to acquire newer cars.

MODEL ASSESSMENT

The model estimation effort yielded coefficient values that are largely reasonable and behaviorally intuitive. This section offers an assessment of the model from several different perspectives, including the significance of the spatial dependency parameter, the goodness of fit of the model relative to a model that does not include spatial dependence, and differences in elasticity estimates between the multinomial probit model with spatial dependency and the independent multinomial probit model that ignores spatial dependency.

Among the weight matrix specifications that were tested, the specification based on inverse distance offered the best fit. The spatial autoregressive parameter in the spatial lag formulation δ turns out to be statistically significant with a value of 0.1872 and *t*-statistic of 3.80. This is evidence of the presence of spatial spillover effects arising because of either didactic interactions of individuals in proximally located households or residential self-selection effects that can lead to a clustering of households with similar preferences in vehicle type choice.

Although the spatial parameter is statistically significant, suggesting superior data fit in the spatial model compared with a corresponding nonspatial model, an alternative way to compare these nested models is through the adjusted composite likelihood ratio test (*8*). The composite log likelihood value for the nonspatial model is −138971.8 (52 parameters estimated) and that for the final spatial model is −138827.6 (53 parameters estimated). The adjusted composite likelihood ratio test statistic of comparison between the two models is 7.84, which is greater than the critical chi-square value of 3.84 associated with one degree of freedom; this demonstrates the presence of spatial interactions in vehicle type decisions. Ignoring such spatial interaction effects results in a model with poorer statistical goodness of fit.

A question often raised in the context of advanced choice models that incorporate additional (observed or unobserved) effects is the extent to which policy forecasts might differ if such effects are ignored. Although the goodness of fit is significantly better and the spatial interactions parameter is significant, would policy forecasts be different if one model were used instead of another (one that ignores spatial dependence effects)? The parameter estimates in Table 1 do not directly indicate the magnitude of the impact of variables on the probabilities of acquiring each vehicle type. To address the question, it is useful to compute aggregate-level elasticity effects of variables for the different model specifications. Specifically, the effects of variables on the expected share of each vehicle type alternative are examined in this paper. This is achieved through computation of the marginal probability that each household will choose a certain vehicle type in a single synthetic choice scenario, followed by aggregation of these probabilities across households and all choice occasions for each vehicle type alternative.

The following procedure is used to compute the shares of each vehicle type alternative. The utility function of vehicle type i for household q is as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} \sum_{t''=1}^{T_{q'}} U_{q't'i} + \boldsymbol{b}'\boldsymbol{x}_{qti} + \tilde{\varepsilon}_{qti} \tag{8}$$

where the notation is similar to that described in the methodology section of this paper. Then, other notation described previously can be used to write

$$\mathbf{U} = \mathbf{S}[\mathbf{xb} + \tilde{\boldsymbol{\varepsilon}}] \tag{9}$$

The preceding $RI \times 1$-vector $\mathbf{U}$ is simulated 500 times with the estimated values of $\mathbf{b}$ and by randomly drawing 500 times from the appropriate normal distributions for $\tilde{\varepsilon}$. Next, the chosen alternative is determined as the alternative with the highest utility for each of the 500 draws. Finally, the predicted share of each alternative across the 500 draws is taken as an estimate of the probability of each vehicle type alternative. The aggregate share (across all households and all synthetic choice occasions) of each alternative is obtained by aggregating the synthetic choice occasion level probabilities of each vehicle type alternative across all households.

The elasticity computed is a measure of the percent change in the aggregate share of each vehicle type alternative caused by a change in an exogenous variable. For dummy variables, the value of the variable is changed to 1 for the subsample of observations for which the variable takes a value of 0 and to 0 for the subsample of observations for which the variable takes a value of 1. The shifts in expected aggregate shares in the two subsamples are then added after the sign of the shifts is reversed in the second subsample, which yields the effective percent change in the expected shares across all households in the sample caused by a change in the dummy variable from 0 to 1. For continuous variables, the value of the variable is increased by 25% for each observation and the percent change in the expected shares is computed. For variables that take only integer values (such as number of full-time workers), the value is increased by unity.

Elasticity estimates are computed for the nonspatial MNP and the spatial MNP model and are presented in Table 2. The first entry in Table 2 indicates that according to the MNP model with no spatial interaction, households with a highest education attainment of a bachelor's degree are 3.43% more likely than other households not to acquire a vehicle at any given choice occasion. Other elasticity effects can be interpreted similarly.

All the elasticity effects are consistent with the parameter estimates in Table 1. Also, the elasticity effects of the spatial and nonspatial models are in the same direction (sign) for all variables. However, the elasticity estimates of the nonspatial MNP model and spatial MNP models are quite different in magnitude. In general, the elasticity effects of the spatial model are consistently higher in magnitude than those from the nonspatial model. For instance, the elasticity effect of the number of full-time workers for the new noncar alternative is 312%, whereas the corresponding number according to the nonspatial MNP model is only 80%. Similarly, the spatial model implies that Hispanic households are 198% more likely to obtain an old noncar, whereas the nonspatial model implies only

TABLE 2 Aggregate-Level Elasticity Effects of Spatial and Nonspatial Models

Variable	No Vehicle		Old Car		New Car		Old Noncar		New Noncar	
	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial	Spatial
Demographics										
Highest education attainment in household (base is college degree or less)										
Bachelor's degree	3.43	2.90	5.20	4.05	8.53	7.34	5.58	4.97	−83.27	−69.29
Graduate	2.71	2.41	4.06	3.50	6.97	5.98	4.61	4.08	−66.52	−57.63
Hispanic status (base category is non-Hispanic): Hispanic origin	−5.85	−14.51	−17.21	−39.66	−15.65	−36.84	84.69	197.79	−14.18	−31.90
Race (base category is all other races): African American	8.44	11.67	−45.50	−63.80	22.52	36.01	22.83	35.76	−25.34	−50.94
Housing tenure (base category is rental home): own	−29.85	−35.61	1.94	7.49	45.89	60.13	50.88	59.74	101.92	91.74
Household income (base category is all lower-income levels): greater than $75,000	−39.03	−43.62	56.80	70.00	45.09	40.81	11.23	11.81	88.77	98.06
Number of adults	−23.62	−68.07	34.20	80.83	15.90	20.32	69.54	234.10	−45.62	−88.07
Number of full-time workers	−11.24	−45.83	−19.58	−65.85	66.15	240.55	−20.07	−66.73	79.67	312.38
Number of people with more than one job	4.49	7.65	10.02	18.13	−52.21	−92.91	9.83	17.08	13.23	24.81
Mean distance to work (mi)	−0.26	−0.59	−0.31	−0.78	−0.71	−1.46	−0.38	−0.77	6.13	13.38
Presence of children										
6–10 years	3.82	6.99	−26.48	−50.18	9.96	20.24	11.09	20.64	7.77	14.70
11–15 years	−1.96	−5.62	−5.77	−16.28	−5.36	−13.92	28.55	77.49	−4.87	−11.79
Presence of senior adults	9.07	8.62	22.45	21.88	23.76	25.47	−72.84	−68.95	−78.65	−79.93
Presence of individual with prolonged medical condition (>5 years)	4.55	7.07	10.14	16.62	−52.70	−85.69	9.88	15.62	13.23	23.29
Immigration status (base is nonimmigrant household): immigrant household	16.82	17.65	−28.88	−31.14	−33.84	−36.13	2.16	3.25	−5.46	−4.04
Existing Vehicle Fleet Characteristics										
Number of old cars	45.81	81.34	−60.05	−98.99	−55.64	−97.71	−50.85	−97.31	−46.11	−93.50
Number of new cars	60.12	81.83	−91.74	−100.00	−57.61	−99.71	−93.96	−100.00	4.57	−85.41
Number of old noncars	39.99	75.88	−50.98	−97.64	−32.43	−86.81	−80.39	−99.93	−8.86	−60.21
Number of new noncars	61.30	83.14	−75.85	−99.98	−52.10	−99.26	−86.40	−100.00	−89.53	−99.97
Accessibility Measures										
Primary arterial roads roadway length within 10 min (mi)	1.81	4.34	4.89	13.70	4.56	12.41	−25.15	−64.82	4.38	11.60
Minor arterial roads roadway length within 10 min (mi)	1.18	2.90	3.26	8.89	2.90	7.82	−16.35	−42.05	2.70	7.15
Collector roads roadway length within 10 min (mi)	−4.51	−16.76	−10.89	−38.14	−10.52	−34.42	58.78	203.70	−9.49	−31.05
Total manufacturing employment that can be reached within 10 min	0.23	0.55	0.45	1.08	−2.44	−6.03	0.42	1.02	0.57	1.60
Total arts employment that can be reached within 10 min	0.65	1.84	−4.41	−12.69	1.83	4.95	1.51	4.92	1.44	3.91

85% higher likelihood for Hispanic households. Although the magnitude of the spatial autoregressive parameter is relatively small, the spatial spillover effect compounds the elasticity estimates because of the circular reinforcing mechanism by which a change in the value of a variable for one household changes utilities of vehicle type alternatives for other nearby households, which in turn causes ripple effects in the utility values of the household for which the variable changed in the first place.

Elasticity estimates differ substantially between a model with and a model without spatial dependency effects. These differences can have dramatic implications for policy forecasts that rely on model parameter estimates to infer the magnitude of behavioral changes in response to a change in input conditions. For instance, Assembly Bill 32, the Global Warming Solutions Act, requires that the state of California reduce its statewide carbon emissions to 1990 levels by 2020 through a combination of transportation and land use planning strategies. One of these strategies is smart land use development in which proximal housing development is encouraged for efficient land use patterns (*10*). The underlying idea is that high-density neighborhoods are less conducive to auto use and thus reduce greenhouse gas emissions. However, according to the results of the present study, the proximal housing development strategy could be far more effective if coupled with transportation strategies that encourage purchases of alternative-fuel vehicles. For example, if the local government subsidizes purchases of alternative-fuel vehicles by households in a new dense housing development, because of social interactions among proximal households other households in the region would be likely to buy similar vehicles. This spatial spillover effect coupled with built-environment characteristics that restrain auto use could reduce emissions substantially. The full potential of such coordinated land use and transportation planning strategies can be evaluated only with models that account for spatial dependencies of proximal decision-making agents, such as the model developed in this paper.

CONCLUSIONS

This paper presented a multinomial probit model of vehicle ownership by type (fleet composition) that explicitly incorporates spatial interaction effects caused by observed and unobserved factors. The model was estimated on the Los Angeles region subsample of the California add-on data set of the 2009 National Household Travel Survey, which includes many accessibility and land use variables that are critical for modeling of vehicle ownership. Underlying the model is a behavioral framework that considers the household vehicle fleet as being constructed over a series of purchase choice occasions; this allows the vehicle fleet size to be endogenously determined while history dependency is simultaneously incorporated in the choice model. In other words, the vehicle type that is acquired at any choice occasion is dependent on the existing household vehicle fleet comprising vehicles acquired at earlier choice occasions. The model considered five choice alternatives for each occasion: two body types (car and noncar), two vintage types (less than or equal to 5 years old and greater than 5 years old), and the choice of acquiring no vehicle. The MACML estimation procedure was used to overcome computational intractability associated with traditional simulation and Bayesian model estimation procedures.

Model estimation results show that many individual and household variables, along with accessibility and land use variables, significantly affect choice for acquiring different vehicle types. More important, it was found that the spatial interaction parameter is statistically significant and the model that incorporates spatial spillover effects offers a superior statistical goodness of fit than a multinomial probit model that does not incorporate spatial dependency effects. It was found that a distance-based spatial interaction function offers the best fit, and the interaction between households drops off as the distance between households increases. A comparison of elasticity estimates offered by the spatial effects choice model estimated in this paper against those offered by a model with no spatial effects showed that elasticity estimates differ substantially when spatial effects are incorporated. The elasticity estimates from the spatial effects model were consistently higher in magnitude, suggesting that interaction effects amplify the extent to which households change behavior in response to changes in input conditions. Incorporation of spatial effects in models of discrete choice behavior can result in substantially different policy forecasts, which has clear implications for transportation planning and policy. Future research efforts could further explore the use of alternative spatial interaction functions, examine whether the findings hold true in other geographical contexts and data sets, investigate model parameter transferability, and attempt to separate unobserved and observed spatial interaction effects by jointly modeling residential location choice with vehicle ownership and fleet composition choice.

ACKNOWLEDGMENTS

The authors acknowledge the helpful comments of four anonymous reviewers on an earlier version of the paper. The authors thank Lisa Macias for her help in formatting this document.

REFERENCES

1. *Sources of Greenhouse Gas Emissions: Transportation Sector Emissions.* U.S. Environmental Protection Agency, 2010.
2. Anselin, L. Thirty Years of Spatial Econometrics. *Papers in Regional Science,* Vol. 89, No. 1, 2010, pp. 3–25.
3. Bhat, C. R., I. N. Sener, and N. Eluru. A Flexible Spatially Dependent Discrete Choice Model: Formulation and Application to Teenagers' Weekday Recreational Activity Participation. *Transportation Research Part B,* Vol. 44, No. 8–9, 2010, pp. 903–921.
4. Mohammadian, A., M. Haider, and P. S. Kanaroglou. Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.
5. Adjemian, M. K., C. Lin, and J. Williams. Estimating Spatial Interdependence in Automobile Type Choice with Survey Data. *Transportation Research Part A,* Vol. 44, No. 9, 2010, pp. 661–675.
6. Anselin, L. Spatial Externalities, Spatial Multipliers and Spatial Econometrics. *International Regional Science Review,* Vol. 26, No. 2, 2003, pp. 153–166.
7. Chen, Y., S. Ravulaparthy, K. Deutsch, P. Dalal, S. Y. Yoon, T. Lei, K. G. Goulias, R. M. Pendyala, C. R. Bhat, and H.-H. Hu. Development of Indicators of Opportunity-Based Accessibility. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2255,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 58–68.
8. Bhat, C. R. The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. *Transportation Research Part B,* Vol. 45, No. 7, 2011, pp. 923–939.
9. Godambe, V. P. An Optimum Property of Regular Maximum Likelihood Estimation. *Annals of Mathematical Statistics,* Vol. 31, No. 4, 1960, pp. 1208–1211.
10. *Climate Action Team Report to Governor Schwarzenegger and the Legislature.* California Climate Action Team, California Environmental Protection Agency, Sacramento, 2006.

The Transportation Demand Forecasting Committee peer-reviewed this paper.