
Volume 44A, issue 9, November 2010
ISSN 0965-8564
ELSEVIER
TRANSPORTATION
RESEARCH
AN INTERNATIONAL JOURNAL
Part A: Policy and Practice
Editor-in-Chief
Debbie A. Niemeier

Transportation Research Part A 44 (2010) 661–675

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Estimating spatial interdependence in automobile type choice with survey data

Michael K. Adjemian [a,*], C.-Y. Cynthia Lin [b,1], Jeffrey Williams [b,2]

[a] USDA Economic Research Service, DC, USA
[b] Department of Agricultural and Resource Economics, University of California, Davis, USA

ARTICLE INFO

Article history:
Received 17 August 2009
Accepted 3 June 2010

Keywords:
Vehicle choice modeling
Spatial autocorrelation
Spatial econometrics
Census data
Survey data
Discrete choice

ABSTRACT

In this article, we show that vehicle type ownership is spatially dependent at both the regional and household-level even after controlling for income and population density. We discuss reasons for the existence of spatial effects in vehicle ownership, and note potential implications for policymakers. Our results point to the importance of spatial relationships in transportation research and highlight the hazards of ignoring their role in affecting transportation outcomes. For example, if vehicle type choice is affected by neighborhood spillovers, agencies that regulate traffic flow and road safety could tailor their choice projections and policy tools to account for such interdependence.

Published by Elsevier Ltd.

1. Introduction

The composition of the surrounding fleet may play a role in the process by which households choose vehicles. Presumably, neighbors display more similar ownership characteristics than more distant drivers. The choices of nearby households may signal auto reliability, normalize perceptions in the case of a new body type like SUV's, or stimulate a desire for conformity. Such spatial interdependence is signified by the presence of spatial autocorrelation (Cliff and Ord, 1973). Autocorrelation alone does not prove that household utility is affected by the actions of their neighbors; alternatively, people with similar preferences may self-select for certain regions. Still, if vehicle type ownership is spatially autocorrelated, aggregate level models of vehicle ownership must take account of spatial effects.

Conventional choice models express household utility as a function of own characteristics and the traits of available alternatives, but do not allow for the possibility of inter-household interaction. Consequently, if social interdependence affects vehicle choice, and if it is correlated with established regressors, research that neglects to weigh its importance may produce invalid estimates. Covariates for remaining regressors will be biased to the degree that they pick up the relationship between the outcome and the missing spatial predictor. Closely related to the issue of spatial interdependence is the appropriate level of aggregation. In the extreme, if every household in the neighborhood has the same characteristics and buys the same type of vehicle, whether out of self-selection or out of herd-like behavior, the analysis of household-level decisions offers no more insight into vehicle choice than the analysis of behavior at the aggregate level.

Using the 2000 San Francisco Bay Area Travel survey (BATS) data, we consider whether households display spatial similarity in their automobile type choice in the nine-county region, and search for evidence of consumer choice interaction. To that end,

* Corresponding author. Address: 1800 M Street NW, Washington, DC 20036, USA. Tel.: +1 202 694 5576.
E-mail addresses: madjemian@ers.usda.gov (M.K. Adjemian), cclin@primal.ucdavis.edu (C.-Y. Cynthia Lin), williams@primal.ucdavis.edu (J. Williams).
[1] Address: 3104 Social Sciences & Humanities, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA. Tel.: +1 530 752 0824.
[2] Address: 2144 Social Sciences & Humanities, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA. Tel.: +1 530 754 7625.

0965-8564/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.tra.2010.06.001

we apply diagnostic tools to reveal spatial autocorrelation in ownership data aggregated to the census tract level, and test concentrations to determine whether they are substantive. We develop and apply a model to evaluate whether spatial effects play a role in household-level vehicle choice, and compare its results to more aggregate analysis at the census tract level.

2. Overview and related work

Over the last few decades, social scientists have devoted a growing interest to the nature and impact of spatial interaction (Anselin and Bera, 1998). Furthermore, the concurrent development of spatial econometric methods and advancements in computer power facilitates the empirical study of applied spatial questions. Although one benefit of formally accounting for spatial effects is to more thoroughly evaluate traditional choice problems, doing so may be central to understand and properly estimate the data generating process. Anselin and Griffith (1988) show that if spatial effects are ignored, data analysis can lead to incorrect inference.

Recently, researchers have found that transportation behavior exhibits signs of spatial interdependence, sometimes termed herding or bandwagon effects. Notable applications include forecasting household travel activity (Scott and Kanaroglou, 2002; Vovsha et al., 2004), modeling the decision to telecommute (Paez and Scott, 2007), and explaining commodity flows on a highway network (LeSage and Polasek, 2005). Further, research into travel mode choice has found that neighborhood behavior features into the household-level decision. Dugundji and Walker (2005) consider whether an individual is more likely to choose a given travel mode when accounting for the decisions of others located in his residential zone. Goetzke (2008) finds that the spatial proximity among individuals affects their likelihood of exhibiting herd behavior in selecting public transit to work. Our aim is to explore whether individuals interested in purchasing a car gather useful information by observing the composition of local ownership, and factor the impact of social networks into their own auto choice.

In addition to the number of cars on the road, the degree of vehicle heterogeneity affects roadway congestion and accident rates. Choice models are commonly used to estimate the parameters relevant to the selection of automobile variety, otherwise known as vehicle type choice. Because the set of car types is categorical, a discrete choice framework is generally applied. Inclusion of a spatial component can, however, make discrete choice modeling difficult, unless certain assumptions are made. For example, Goetzke's (2008) spatial lag term is assumed to be exogenous, and he also assumes no spatial autocorrelation in the error term of his mode choice utility specification. As a result, he is able to estimate the significance of the spatial parameter using standard discrete model statistical software. Mohammadian et al. (2005), also make the simplifying assumption of an independent error term in their spatial logit specification of a residential choice model, and use simulated likelihood methods to recover relevant parameters with a mixed logit model. Following this methodology, we condition household utility on observed auto choices, and model spatial effects exogenously.

3. Data

Household vehicle fleet ownership and socioeconomic/demographic data for 15,064 households were collected by the 2000 San Francisco Bay Area Travel Survey (BATS), and authorized by the Bay Area Metropolitan Transportation Commission (MTC). Between February 2000 and March 2001, BATS was conducted in the nine counties that make up the region. However, the residential addresses of survey participants are not reported. Instead, BATS geocoded the location of each surveyed household, and associated every home with its pertinent census tract. The survey achieved a 99.9% success rate in linking the home addresses of surveyed households in this manner. For our analysis, we subsumed the location of surveyed households to the census tract geographical centroid. Distances between tracts are calculated using the Haversine function with the latitude and longitude coordinate inputs listed by the US Census Bureau.

For the aggregate analysis, proportionate auto ownership for each tract is calculated by averaging over the vehicle types exhibited by its surveyed households. Explanatory variables are drawn from census information imported from the year 2000 United States Census Summary File 3. According to the US Census Bureau, tracts contain 4000 people on average, and are specifically designed to group relatively homogeneous individuals in terms of demographics and economic status (US Census Bureau, 1994). The census tracts in the San Francisco Bay Area cover about six square miles of land area, on average. Additionally, census tracts are intended to be permanent statistical subdivisions, increasing their usefulness in empirical applications. Census measures include mean household size, racial composition, average age, average educational attainment, marital status, and median income.

We classified the vehicles in BATS into nine types according to those used by the auto information company Edmunds.com: coupe, compact sedan, mid-size sedan, large sedan, station wagon, sports utility vehicle (SUV), pickup truck, minivan/van, and sportscar. We created two indicator variables: whether the vehicle was new at the time of the study (model year 2000), and whether it was made by a premium automaker, such as BMW, Ferrari, or Porsche. Therefore, we investigated the presence of spatial interdependence in the ownership of eleven categories of car ownership.

Those tracts that did not display sufficient observations according to the definition of proportionate ownership are excluded from the aggregate analysis. Out of the 1332 tracts surveyed by BATS, requiring that a tract contain at least 10 surveyed homes limited the sample to 560 Bay Area tracts. We excluded those cars in BATS that could not be identified or classified. After removing the sparsely sampled tracts and those cars that could not be identified, we used a sample of 8837 cars in the aggregate analysis. The average tract in the sample contains 15 surveyed homes.

For the disaggregate analysis, the dependent variable is the binomial outcome associated with the purchase of a given body type, conditional on the decision to buy a new car. Census information for block group density and median housing age is available in the year 2000 United States Census Summary File 3. The remainder of the explanatory variables are imported directly from BATS. About 6%, or 1660, of the original BATS vehicles, represent model year 2000 cars. Of these, 439 do not contain information essential to this study, such as self-reported household income, employment status, or age. We define the neighborhood for each household as the five-mile radius around its location. In order to produce meaningful estimates, we further eliminated 64 observations that did not contain at least 30 neighbors within this threshold. After cleaning the data, 1157 BATS vehicles are represented in our disaggregate data.[3]

4. Aggregate methods

Statistical inference from models that do not account for existing spatial clustering suffer from a loss of efficiency, since an independent sample of the same size contains more information, and may produce biased and inconsistent estimates. Although improving the sampling scheme may suffice, models themselves can be modified to control for the spatial dimension of the data. Our reliance on a previously conducted survey precludes using corrective measures like increasing the sample size, so we instead incorporate spatial components and test for their significance in our regression analyses. In the following sections, we discuss the methods used to determine whether aggregate level transportation choice exhibits spatial autocorrelation, review models that relax the assumption of spatial independence, which are explained in detail by Anselin (1992), and discuss their relevance to the study of vehicle choice.

One way to consider the factors associated with vehicle choice, and transportation behavior more generally, is to observe and analyze the collective actions of consumers, relating aggregate, also termed "ecological" travel behavior data to regional level characteristics. A risk of using ecological data is that the explanatory variables may be sufficiently correlated to risk multicollinearity, which increases the standard errors and confuses the significance of parameter estimates. On the other hand, aggregate level data are often much easier to obtain and computation requirements are less taxing than in a comparable, micro-level framework.

4.1. Detecting spatial autocorrelation

The presence of spatial autocorrelation signifies that a variable is spatially dependent. Positive spatial autocorrelation is discernible as spatial clusters. For example, as shown in Fig. 1, BATS data show that census tracts with like rates of pickup truck ownership are grouped together. Fig. 2 demonstrates a similar pattern for compact cars. Negative spatial autocorrelation is the opposite, resembling the layout of a checkerboard, where high values are surrounded by low values and vice versa.

In order to formally detect whether auto ownership is spatially autocorrelated, we compute the Moran's I statistic, with the Empirical Bayes correction for the variance of rates (Assuncao and Reis, 1999). Significance tests indicate whether automobile type ownership is randomly distributed over the Bay Area. If the null hypothesis of no spatial autocorrelation is rejected, the auto type in question is determined to be spatially dependent, with a positive or negative relationship signified by its sign.

As show in Table 1, in the San Francisco Bay Area, compacts, large sedans, mid-size sedans, pickup trucks, minivans, station wagons, SUVs, premium and new cars are all positively spatially autocorrelated at a p-value lower than 5%. The null hypothesis could not be rejected at the 5% level for coupes, and sportscars. These results indicate the presence of considerable clustering in BATS automobile ownership data. However, the significance of Moran's I does not imply that proportionate ownership is truly spatially dependent (Lin, forthcoming). Quite possibly, other factors might be the source of the spatial autocorrelation. For example, the dependence may no longer be significant after explanatory variables are considered.

4.2. Dependent variable: proportionate auto ownership

For the purposes of this study, aggregate level vehicle choice is defined as the proportion of households in a census region that own a similar car. The regional level considered in our analysis is that of census tract, since it is the lowest level of aggregation that allows a useful number of observations. For a body type j, the vehicle rate y_j is defined by:

$$y_j = \frac{N_{j,t}}{C_t}, \quad \text{for } j = 1, 2, \ldots, 11 \text{ and } t = 1, \ldots, T, \tag{1}$$

where

$$N_{j,t} = \sum I(n_{j,t} = i), \quad \text{for } j = 1, 2, \ldots, 11, \tag{2}$$

and

[3] We confirmed that the means of the excluded observations were very similar to the sample means reported in Table 5.

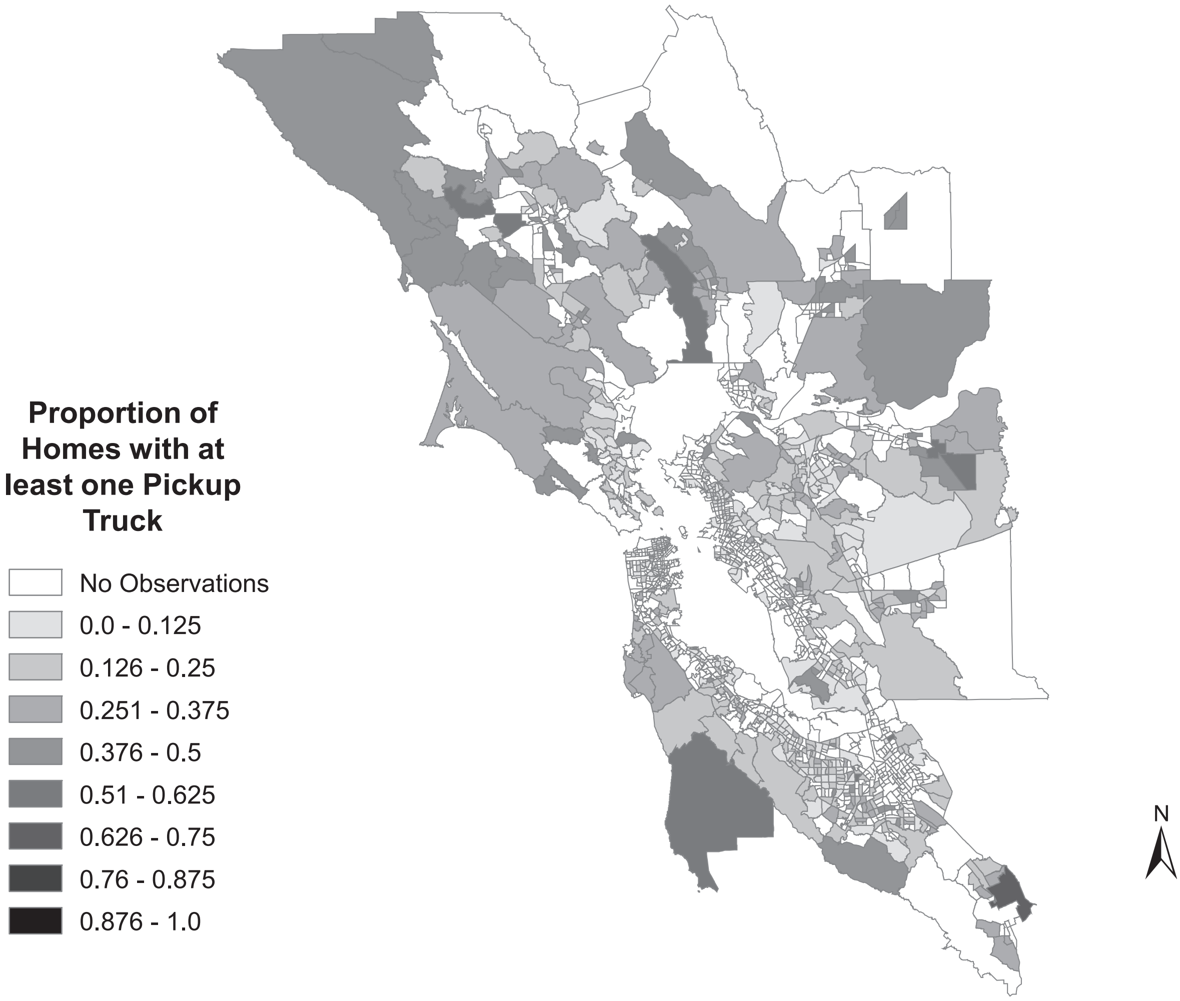

Fig. 1. Pickup truck ownership in San Francisco Bay Area census tracts.

$$C_t = \sum_i n_{j,t}, \quad \text{for } j = 1, 2, \ldots, 9 \tag{3}$$

Here, $j = 1, \ldots, 9$ refers to the nine different body type classifications for the automobiles in BATS: compact sedan, SUV, etc. The other two values for j (10, 11) indicate whether the vehicle in question can also be termed a premium or new car. N, the total number of surveyed households with cars of type j in census tract t, is the sum over the indicator functions, I, which equal one if household $n_{j,t}$ owns a car of type j, and zero otherwise. The total number of surveyed households in the tract is represented by C. Since y_j is a proportion of households, the number of tracts, T, is rationed by setting minimum levels of observations in order to ensure meaningful estimates.

4.3. Naïve aggregate model

We define the naïve model so that it relates proportionate auto ownership of car body type j to the characteristics of the region under consideration, without taking account of spatial effects. For example pickup truck ownership is a function of local variables alone. A general form of the aggregate type ownership model is given by a simple linear model:

$$y = x\beta + \varepsilon \tag{4}$$

where y is a $T \times 1$ vector of dependent variables that represent the rate of auto type i ownership for every region t. Regional characteristics are represented by the $T \times K$ matrix x, while the $K \times 1$ vector β indicates the relationship between the regressors and the outcome, where K is the number of regressors. After reviewing recent type choice models (Choo and Mokhtarian, 2004; Mohammadian and Miller, 2003), we determined that census tract level traits likely to be related to auto type ownership include median income, average age, gender and household size in number of residents. We also included population density, since residents in high density areas may exhibit different preferences than their counterparts in rural communities.

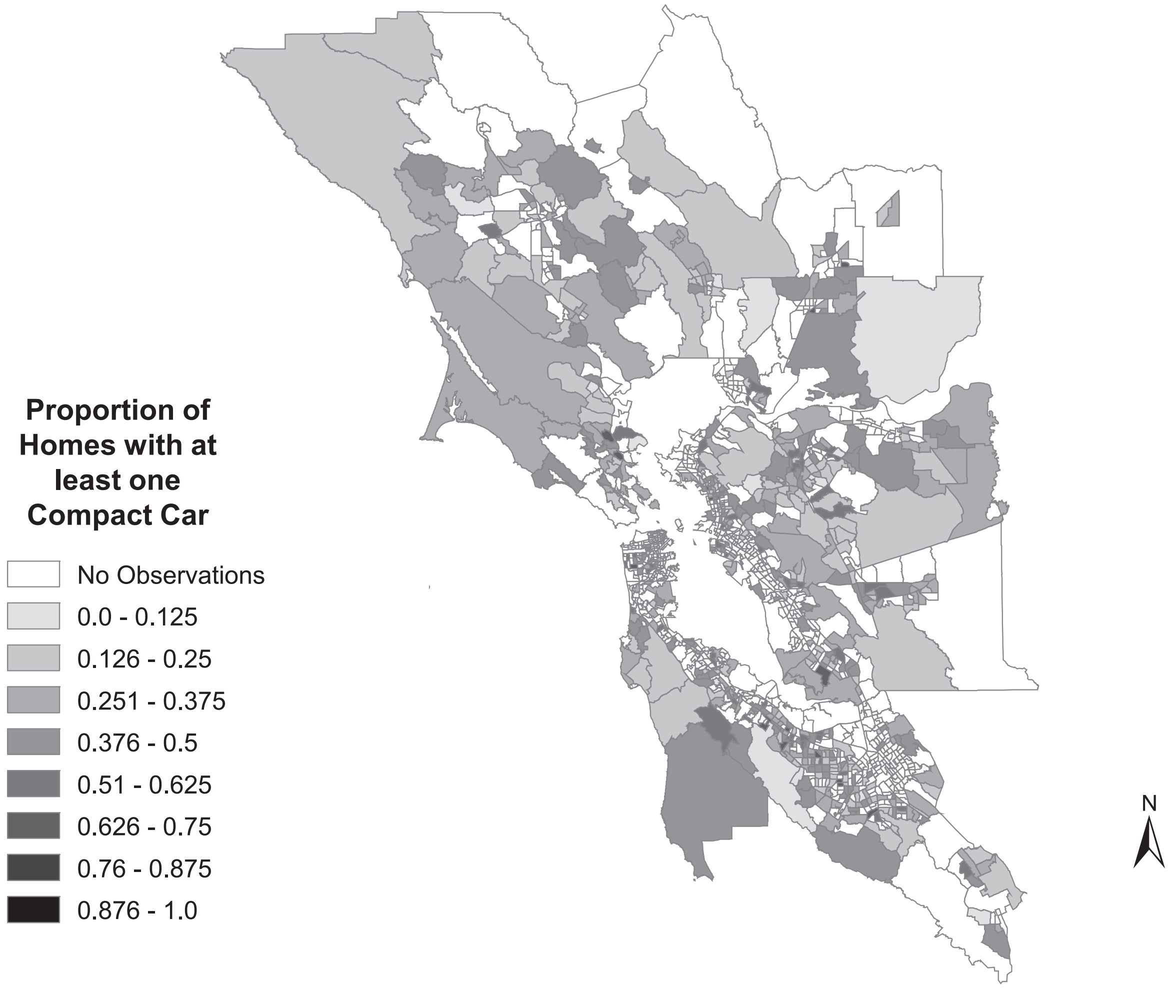

Fig. 2. Compact car ownership in San Francisco Bay Area census tracts.

Table 1
Moran's I[a] and model selection test for aggregate-level ownership of Bay Area vehicles.

Auto type	I	p-value	LM tests	C.F.[b] test
Compact	0.095	0.001	OLS	–
Coupe	−0.029	0.828	OLS	–
Large sedan	0.125	0.001	OLS	–
Mid-size sedan	0.153	0.001	OLS	–
Minivan/van	0.136	0.001	OLS	–
Pickup truck	0.371	0.001	Lag	Lag
Station wagon	0.147	0.001	Lag	Error
SUV	0.155	0.001	Lag	Lag
Sportscar	0.028	0.148	OLS	–
Premium	0.204	0.001	OLS	–
New car	0.090	0.001	OLS	–

[a] Estimated via 1000 Monte Carlo simulations of Empirical Bayes Index for 560 census tracts.
[b] "C.F." stands for "common factor".

The $T \times 1$ error vector ε is usually assumed to be independent and identically distributed, and is implicitly spatially random. If these assumptions are valid, then the naïve model can be estimated by ordinary least squares (OLS). If, however, the error term is spatially correlated, then the OLS assumption of independent error is violated, and its estimates can lead to incorrect inference. Although the coefficients are unbiased as long as $\text{cov}(x, \varepsilon) = 0$, the result is a loss of efficiency, meaning that the statistics representing the significance of regression parameters will be biased, as well as the measure of model fit. Therefore, if spatial dependence is present, uncorrected statistical results can be misleading.

4.4. Aggregate error model

One way to address spatial dependence in the error term is to account for it formally in the model. Ordinarily, the error term from the naïve model is modified to follow an autoregressive process, where the relationship between locations is defined by a weight matrix, W. The error model is then:

$$y = x\beta + \varepsilon, \quad \text{where } \varepsilon = \lambda W\varepsilon + \xi \tag{5}$$

The weight matrix is a $T \times T$ matrix with zeros on the diagonal so that the residual in a particular location cannot affect itself. The remaining values in W indicate the amount of influence each tract location is assumed to have on every other census tract in the dataset. In this paper, we define each element in W to be 1 if tract t is among the "nearest neighbors" to tract k according to the Great Circle distance between the latitude and longitude coordinates of the respective regional centroids, according to the US Census Bureau, and zero otherwise. Now, the dependence in ε is modeled explicitly, and its magnitude is represented by the coefficient λ. The term ξ is assumed to be independently and homoskedastically distributed.

If the spatial error model effectively explains the spatial dependence of the system, then it can recover efficient estimates of β, in contrast to the naïve model, although β is unbiased in either case. Lin (forthcoming) refers to this situation as one of "spurious" spatial dependence, since the efficiency loss from estimation can be avoided once we control for the non-spherical error term. The additional complication of an unknown autoregressive parameter, however, makes OLS less preferable than other approaches to estimation, namely maximum likelihood (ML) or the generalized method of moments (GMM).

4.5. Aggregate lag model

On the other hand, if the true model is one where proportionate ownership in a given tract is mutually influenced by the value of the dependent variable in other tracts, then neither the naïve nor the error model adequately specifies the system. To avoid omitted variable bias, the spillover effect from one tract to another must be controlled for explicitly. The autoregressive aggregate model is then:

$$y = \rho Wy + x\beta + \varepsilon \tag{6}$$

Again, the weight matrix defines the structure of spatial interdependence, so the expression Wy represents the spatially weighted average of nearby auto ownership, with zeros on the diagonal. Assuming the model is correctly specified, the significance of ρ indicates whether proportionate ownership is substantively spatially dependent. A positive ρ represents ownership spillovers, after controlling for the predictors in x. In that case, ownership of a given body type is concentrated after accounting for factors like population density, average age and median income.

Yet, the introduction of a spatial lag makes OLS parameter estimates biased and inconsistent, since $\mathrm{cov}(y_m, \varepsilon_l) \neq 0$ for some m not equal to l, and the dependent variable is correlated with the error term. Instead, the model can be suitably estimated by ML. The choice of the weight matrix is an important question in applied transportation work, particularly when the extent of spatial interaction is difficult to discern (Kawabata and Shen, 2007).

4.6. Choosing the weight matrix

Crucial to both the spatial lag and error models, the weight matrix plays a central role in the estimation of spatially dependent systems. Effectively, the weight matrix defines the neighborhood for each census tract, and enumerates the extent of the interaction between each tract and those nearby. If the data are characterized by contiguity, like the census tracts surveyed by BATS, then neighbors can be determined on the basis of sharing a border. However, since BATS did not evenly sample every tract in the Bay Area, we chose to define a tract's neighbors as the $k = 5$ closest tracts that met minimum observational thresholds, and reported those results.[4] The nearest neighbor approach can be advantageous in situations that evaluate census tracts in both rural and urban areas, such as this study, since it constrains the number of neighbors to equivalence (Anselin, 2002).

4.7. Which model to use?

Although a diagnostic test like Moran's I provides evidence that a variable is spatially autocorrelated, it does not explain why such dependence occurs. Moreover, the presence of spatial autocorrelation does not indicate whether to specify the resulting model in an error, lag, or non-spatial format. This specification is an important step, since it affects that way that we build the model and interpret any spatial dependence in auto ownership. A variety of methods have been proposed to address this problem (Anselin and Bera, 1998), and we present two separate tests in this article.

Anselin et al. (1996) showed that OLS residuals provide a guide for model selection, and introduced a series of Lagrange Multiplier (LM) tests that diagnose the presence of autocorrelated errors and misspecification possibilities such as a missing error process or absent spatial lag. According to Florax and Vlist (2003), the LM tests adequately determine the correct model

[4] We verified that our results are robust to various values for k.

design. Additionally, the LM tests explicitly allow for the possibility that the OLS model describes the system properly, and that no spatial model should be used. Still, they do not offer a direct test between the spatial lag and spatial error model, but have the advantage of simplicity, since the method requires only the estimation of an OLS model.

4.8. Durbin model: testing for "true" spatial dependence

If spatial effects are known to occur and the naïve model is insufficient, a more elegant way to determine the proper specification is to relate the error and lag models using a likelihood ratio test. A potential complication is that the likelihood ratio is only valid in the case of nested models, and this does not immediately apply to the previous sections. Fortunately, it is possible to restate the error model in Eq. (5), beginning by rearranging the error process:

$$\varepsilon = \lambda W\varepsilon + \xi \rightarrow \varepsilon = (I - \lambda W)^{-1}\xi \tag{7}$$

Substituting into Eq. (5), and organizing terms:

$$y = x\beta + (I - \lambda W)^{-1}\xi \rightarrow \tag{8}$$

$$y = \lambda Wy + x\beta - \lambda Wx\beta + \xi \tag{9}$$

Equivalently, (8) is a special case of the lag model, frequently termed the Durbin model, where the explanatory variables are composed of $[x\ Wx]$, so that distanced versions of the ordinary predictors are included. Without specifying a relationship between the parameters, an unconstrained Durbin model is given by:

$$y = \lambda Wy + x\beta - \delta Wx + \xi \tag{10}$$

In the literature, the nonlinear constraint that $\delta = \lambda\beta$ is referred to as the common factor hypothesis. Under the null hypothesis, spatial dependence is adequately specified by an error model, the constraint holds, and the Durbin model in Eq. (9) collapses to the original spatial error specification, Eq. (5). The existence of the common factor is tested by means of a likelihood ratio. Whether or not the null hypothesis is rejected depend on the increase in log likelihood, from model (5)–(9). If the null hypothesis is rejected, it indicates that the error model does not suitably account for the spatial autocorrelation in the dependent variable. In that case, the spatial lag model is preferred.

This method is limited, though, by the fact that the common factor test requires that the lag model be reformulated to include lagged explanatory variables, although these may not belong in the model. Also, as opposed to the LM decision rule, it does not allow for the possibility that the naïve model is adequate. Still, this test compares the error and the lag specifications directly, and provides evidence in favor of "true" spatial dependence if the null hypothesis is rejected.

4.9. Estimation strategy

In order to determine whether proportionate ownership is spatially dependent, we perform LM tests on OLS errors for all eleven car types. If the LM tests suggest that a spatial model is appropriate, we conduct a common factor test to verify robust model selection. Taken together, we select the appropriate model, and estimate it in order to search for spatial dependence, controlling for potential confounders such as income and population density.

5. Aggregate results

5.1. Descriptive statistics

Table 2 displays descriptive statistics for the variables we used to model aggregate vehicle type ownership. Each of the first nine auto types shown stands on its own as a separate body type. Entries are also listed for those vehicles that can be further classified as "premium" or "new cars". Ownership levels do not sum to one, since many households own multiple cars. Compact and mid-size sedans are the most popular vehicle types represented in BATS, owned by 35% and 40% of households, respectively. The least common car types in the Bay Area are station wagons, sportscars, and coupes.

Explanatory variables drawn from US Census tract data are displayed in the table's lower panel, and exhibit considerable diversity over the Bay Area. For example, census tract 207 in downtown San Francisco has over five thousand residents squeezed into just over 1/10th of a square mile of land area, while Sonoma County census tract 1543.01 has about 4000 residents is spread over a 400 square mile area. The considerable range and variation in the explanatory variables avoids potential problems with estimation when using aggregate data (Adjemian, 2009; Adjemian and Williams, 2009).

5.2. Spatial regression

Nearly every auto type in BATS is spatially autocorrelated. We use spatial regression to evaluate whether aggregate auto ownership remains clustered, after controlling for the explanatory variables in Table 2. All regressions are estimated via maximum likelihood. Table 1 shows the results of the model specification tests. Regression results for vehicle types best rep-

Table 2
Descriptive statistics for variables used in aggregate spatial model (minimum 10 households per tract).

Variable	Obs.	Mean	Std. dev.	Min	Max
Ownership					
Compact	560	0.35	0.13	0	0.74
Coupe	560	0.10	0.08	0	0.38
Large sedan	560	0.11	0.09	0	0.55
Mid-size sedan	560	0.40	0.14	0.07	0.92
Minivan/van	560	0.12	0.09	0	0.50
Pickup truck	560	0.19	0.13	0	0.70
Station wagon	560	0.05	0.07	0	0.40
SUV	560	0.19	0.12	0	0.82
Sportscar	560	0.10	0.08	0	0.43
Premium	560	0.17	0.12	0	0.64
New car	560	0.12	0.09	0	0.43
Explanatory variables for census tracts					
Population density (people/sq mi)	560	6478	6371	11	42,538
Log of median income	560	11.19	0.32	9.87	12.04
Proportion w/bachelor degree	560	0.20	0.07	0.04	0.43
Avg. age	560	42.79	4.91	22.48	54.12
Avg. household size	560	2.53	0.43	1.13	4.20
Median yr housing was built	560	1966	13	1939	1995
Proportion of female residents	560	0.51	0.02	0.41	0.67
Proportion of married residents	560	0.46	0.07	0.14	0.60
Proportion of black residents	560	0.03	0.06	0	0.53
Proportion of Asian residents	560	0.15	0.13	0	0.70
Proportion of Latino residents	560	0.12	0.09	0.01	0.55

resented by OLS are displayed in Table 3. Table 4 shows the same for vehicle types that necessitate a spatial coefficient. As the specification tests disagreed in the case of station wagons, we include results of both spatial models.

The regression results in the tables are interpreted as the percentage point increase in the proportion of homes that own a given vehicle type correlated with a one unit—or otherwise specified—increase in the regressor. For example, column 2 in Table 3 reports that a one person increase in the average Bay Area home is associated with a four point decrease in the proportion of homes that own a compact car. The lone exception is the coefficient on the spatial terms. A significant coefficient

Table 3
Aggregate results for vehicle types best represented by OLS.

	Compact	Coupe	Lg sedan	Md sedan	Minivan	Sportscar	Premium	New car
Pop density (+100 K/sq mi)	−0.03	0.08	−0.1	0.2	−0.06	−0.23***	−0.05	−0.03
	(0.14)	(0.09)	(0.09)	(0.14)	(0.09)	(0.08)	(0.11)	(0.09)
Log median income	0.07*	0.02	0.03	0.11***	0.02	0.07***	0.15***	0.05**
	(0.04)	(0.02)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)
Pct. bachelor degrees	0.18	0.02	0.10	0.03	0.18	0.03	0.28**	0.02
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Avg. age (+10 yrs)	−0.03	0.01	0	−0.01	0.07***	−0.01	0.02	−0.01
	(−1.77)	(−1.77)	(−1.77)	(−1.77)	(−1.77)	(−1.77)	(−1.77)	(−1.77)
Avg. HH size (people)	−0.04*	0	0.03*	0.01	0.03*	0.02	−0.02	0
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Median yr house built (+10 yr)	0	0.01	0	0	0	0	0	0.01***
	(0.01)	(0)	(0)	(0.01)	(0)	(0)	(0)	(0)
Pct. female residents	−0.16	−0.14	−0.03	0.71***	0.21	−0.18	0.41*	0.01
	(0.27)	(0.17)	(0.18)	(0.28)	(0.18)	(0.16)	(0.21)	(0.17)
Pct. married	−0.45***	−0.06	0.27**	0.49***	−0.16	−0.14	0.06	0.08
	(0.16)	(0.1)	(0.11)	(0.17)	(0.11)	(0.1)	(0.13)	(0.1)
Pct. black residents	0.14	−0.02	0.1	0.01	0.02	−0.05	0.04	−0.04
	(0.11)	(0.07)	(0.07)	(0.11)	(0.07)	(0.06)	(0.08)	(0.07)
Pct. Asian residents	0.13**	−0.02	−0.09**	−0.07	0.07*	0.08**	−0.08**	0.04
	(0.06)	(0.04)	(0.04)	(0.06)	(0.04)	(0.03)	(0.04)	(0.04)
Pct. Latino residents	0.04	0.05	−0.11*	−0.1	−0.15**	−0.02	0.02	0.09
	(0.09)	(0.06)	(0.06)	(0.09)	(0.06)	(0.06)	(0.07)	(0.06)
Constant	−0.81	−1.07	0.25	−0.77	−0.27	−1.14*	−2.48***	−2.21***
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	560	560	560	560	560	560	560	560

Std. errors rounded to two decimals in parentheses.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 4
Aggregate results for vehicle types best represented by spatial models.

	Pickup		Statnwgn			SUV	
	OLS	Lag	OLS	Lag	Error	OLS	Lag
Spatial effects (lagged ownership/error)		0.18***		.217***	.223***		0.1*
		(0.05)		(0.06)	(0.06)		(0.06)
Pop density (+100 K/sq mi)	−0.28**	−0.27**	−0.14**	−0.1	−0.08	−0.01	−0.01
	(0.12)	(0.12)	(0.07)	(0.07)	(0.07)	(0.12)	(0.12)
Log median income	−0.09***	−0.07**	−0.03	−0.03	−0.03	0	−0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Pct. bachelor degrees	−0.46***	−0.4***	0.09	0.08	0.1	0.47***	0.47***
	(0.14)	(0.14)	(0.09)	(0.08)	(0.09)	(0.15)	(0.15)
Avg. age (+10 yrs)	−0.01	−0.01	0.01	0.01	0.01	0.02	0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Avg. HH size (people)	0.07***	0.06***	0.01	0.01	0.01	0.08***	0.08***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Median yr house built (+10 yr)	0	−0.01	−0.01***	−0.01***	−0.01***	0.01***	0.01**
	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pct. female residents	−0.81***	−0.76***	0.08	0.08	0.1	−0.25	−0.25
	(0.23)	(0.22)	(0.14)	(0.13)	(0.14)	(0.23)	(0.23)
Pct. married	0.17	0.13	0	0.01	0.02	0.07	0.09
	(0.14)	(0.13)	(0.08)	(0.08)	(0.08)	(0.14)	(0.14)
Pct. black residents	−0.06	−0.05	0.02	0.01	0.03	−0.17*	−0.15*
	(0.09)	(0.09)	(0.05)	(0.05)	(0.06)	(0.09)	(0.09)
Pct. Asian residents	−0.27***	−0.24***	−0.02	−0.01	−0.02	−0.16***	−0.15***
	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Pct. Latino residents	−0.03	−0.01	−0.08*	−0.07	−0.08*	−0.03	−0.04
	(0.08)	(0.08)	(0.05)	(0.04)	(0.05)	(0.08)	(0.08)
Constant	2.21**	2.61***	2.09***	1.79***	1.86***	−2.32**	−2.02**
	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	560			560		560	
AIC	−893	−902	−1468	−1481	−1480	−866	−867

Std. errors rounded to two decimals in parentheses.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

on the spatial lag is a sign that auto ownership for that body type remains spatially correlated after controlling for the other explanatory variables. In this case, since the spatial lags are all positive and significant, this means that vehicle type ownership is concentrated spatially, and may be characterized by spillovers. A significant autoregressive error term indicates that unobservables are correlated over space.

Compared to traditional OLS, the spatial approach to modeling aggregate vehicle type choice offers two important advantages. First, and most importantly, since OLS may produce inconsistent and/or inefficient estimates, it cannot necessarily be relied upon for correct inference. Evidence of such a problem is apparent in Table 4, where the OLS model for station wagon ownership assigns significance to census tract population density. After accounting for the spatial filter, the coefficient on population density is no longer statistically important.

Another advantage offered by the spatial models is that they offer a better data fit than do the conventional models. Because it provides an additional parameter, spatial models inevitably produce a likelihood gain. For each body type in Table 4, we compared Akaike's Information Criterion (AIC) values to determine if this gain is sufficient to outweigh the penalty for the lost degree of freedom. In every case, spatial models produced a lower AIC, and thus represent the preferred method to model vehicle ownership. Even for station wagons, although specification tests disagreed, the lag model carries a slightly better AIC.

6. The ecological fallacy

Aggregate level results describe the collective behavior of consumers. For example, a model that estimates vehicle choice using explanatory variables drawn from local US Census tracts can adequately explain choices at the tract level. However, Robinson (1950) showed that ecological, or aggregate level, relationships do not necessarily translate to the individual level. Unless homogeneity constraints are imposed, it would be premature to infer household-level behavior from aggregate data. Still, the existence of persistent, clustered vehicle type ownership among census tracts indicates both the importance of spatial effects, and the possibility that conventional transportation choice models should perhaps be modified accordingly. We use BATS survey data to empirically test whether household-level vehicle choice exhibits spatial effects.

7. Disaggregate methods

Another modeling approach is to consider the household as the decision making entity, linking micro-level characteristics to choice outcomes. Household-level models have the advantage of better capturing consumer behavior, and as a result, the relationship between household characteristics and choice outcomes. For this reason, they may be more useful for policy analysis (Zhao and Kockelman, 2002). Although the data requirements are more exhaustive, disaggregate models are the preferred method of modeling vehicle choice (Bhat and Pulugurta, 1998).

The level of a given household's automobile ownership affects its propensity to select a transportation mode, its destination of interest in leisure activities, and the number of trips it makes (Nobile et al., 1997). Disaggregate models focused on predicting the number of cars chosen by a household are used to provide inputs into transport projection models (De Jong et al., 2004). Likewise, car type choice models are used to project fleet composition, an important component of models used to predict non-point source pollution and road network congestion. Given that car ownership is a categorical variable, and since type choice is made among a known set of possibilities, econometric methods used for parameter estimation are almost exclusively discrete choice models.

7.1. Household spatial lag model

Perhaps the simplest way to conceive of auto type choice at the disaggregate level is to consider vehicle choice as a series of binary choices among vehicle types. This kind of reduced-form model can reveal what explanatory variables correlate with a given household's likelihood of selecting a certain auto body type, and provides some indication as to whether the decision is influenced by spillovers. For example, Goetzke (2008) uses a binary model to consider whether New York City residents exhibit spatial interdependence in the decision to take public transit to work, even though such consumers face a wide variety of travel alternatives.

Due to data constraints, our model is conditional on the decision to purchase a new car. We model only the demand side of the auto market, and assume that the supply of new cars is perfectly elastic. The logit model is a natural candidate, since it applies random utility theory to the type ownership decision: observed household decisions reveal subjective valuation on the part of consumers.

In this model, household i chooses an auto body type in order to maximize its own utility. For each automobile type j, let the utility for an individual household be given by

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{11}$$

where V_{ij} represents the deterministic portion of utility, and ε_{ij} denotes a random component. Traditionally, deterministic utility is then defined as being composed of a vector of explanatory variables multiplied by parameters:

$$V_{ij} = \beta' x_i \tag{12}$$

where x_{ih} represents the relevant characteristics of household i, such as income, householder age, local population density, and other demographic variables. Coefficients on explanatory variables indicate the degree to which they affect individual utility. We modify the model to account for the possibility that consumer interaction affects household utility, by including a spatial term

$$V_{ij} = \beta' x_i + \rho' W f(V_{ij}) \tag{13}$$

Here, W is the spatial weight matrix that defines the neighborhood for every household i. We specified W so that all type choices within a five-mile radius of i are given the same weight, and f so that all automobiles are considered by i. As a result, the spatial lag variable is calculated as the proportion of vehicles within a five-mile radius around household i that are similar to the car type being modeled. For example, for the minivan model, the spatial lag variable $Wf(V_{ij})$ would represent the proportion of total cars within a five-mile radius around i that are minivans. Any spatial effect is captured by the parameter ρ.

The household chooses a car body type—the dependent variable y—in the choice set in order to maximize utility. The body type being considered, j, is equal to 1. Conversely, k represents the alternative, i.e., not j. More specifically, since the model is conditional on the decision to purchase a new car, a household that chooses not j is in fact choosing to buy another unspecified auto type. The decision rule for the household is expressed as

$$\begin{aligned}
\Pr[y=j] &= \Pr[U_{ij} > U_{ik}] \quad \forall k \neq j \\
&= \Pr[U_{ik} - U_{ij} < 0] \\
&= \Pr[V_{ik} + \varepsilon_{ik} - V_{ij} - \varepsilon_{ij} < 0] \\
&= \Pr[\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}] \\
&= \int I(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}) f(\varepsilon_{ik}) d\varepsilon_{ik}
\end{aligned} \tag{14}$$

Here, the indicator function I takes a value of one if the expression in parentheses is true, and zero otherwise. We maintain the assumption of independent random error. Additionally, we assume that ε is identically, Bernoulli distributed for all households. Analytically, it can be shown that the probability that household i decides to own body type j is the familiar logistic probability given by

$$P_{ij} = \frac{1}{1 + \exp(V_{ij})} \quad (15)$$

Although we make the simplifying assumption of no spatial autocorrelation in the error term, estimation can be complicated by the fact that the spatial expression in Eq. (12) may pose an endogeneity problem. One way to think of the problem is that the spatial spillover may be multi-directional. For example, if household i's choice affects household k's choice, perhaps household k's choice also influences household i's choice. To avoid this obstacle, Goetzke (2008) makes the assumption that the spatial effects in public transport decisions are exogenously determined.

We circumvent the potential endogeneity by appealing to two explanations: the nature of car purchases, and the temporal indicators in the BATS data. Transactions costs constrain a household's car purchases to be fixed in the short term. As a result, spatial spillovers in auto choices are necessarily unidirectional. In addition, the BATS data denotes each car's month and year of purchase, allowing us to condition the disaggregate model on observed local auto type choices.

8. Disaggregate results

8.1. Descriptive statistics

The mean, standard deviation and range of values for explanatory variables used in the disaggregate model are shown in Table 5. Auto ownership averages about 2 cars, while 74% of the sample records represent home owners. Multicollinearity is not a serious concern in the data, since the highest variance inflation factor (VIF) is 2.1 (Mendenhall and Sincich, 1996). The range and standard deviation of the explanatory variables provides ample variation. Half of the householders are female. Minorities make up around one in five of surveyed homes. Minimum car ownership and occupancy per household is one, by survey construction. However, two homes contained seven members and one surveyed home owned as many as eight automobiles. The mean householder age of 45 is bracketed by a minimum age of 18, and a maximum of 88. The mean lagged ownership variables represent the proportion of like vehicles to the one modeled within a five-mile radius around the typical surveyed home.

Table 5
Descriptive stats for variables used in disaggregate model.

Variable	Obs.	Mean	Std. dev.	Min	Max
Census block group characteristics					
Pop density (pop/sq mi)	1157	7816	9066	16	172,898
Median yr house built	1157	1967	15	1939	1999
Household characteristics					
1999 Log income	1157	11.38	0.53	8.52	12.01
Household vehicles	1157	2.14	0.84	1	8
Household size (people)	1157	2.56	1.21	1	7
Household owned	1157	0.74	0.44	0	1
Licensed drivers	1157	1.88	0.56	0	5
Householder female	1157	0.50	0.50	0	1
Householder age	1157	45	13	18	88
Latino householder	1157	0.04	0.19	0	1
Black householder	1157	0.02	0.13	0	1
Asian householder	1157	0.13	0.33	0	1
Job: agriculture/forestry	1157	0.01	0.11	0	1
Job: construction	1157	0.06	0.23	0	1
Job: HVAC/service	1157	0.01	0.07	0	1
Job: transportation	1157	0.06	0.23	0	1
Job: legal	1157	0.02	0.14	0	1
Job: finance/insurance/R.E.	1157	0.11	0.31	0	1
Job: auto service	1157	0.03	0.16	0	1
Lagged ownership					
Compact	1157	0.23	0.04	0.06	0.33
Coupe	1157	0.06	0.01	0.01	0.17
Large sedan	1157	0.07	0.02	0.02	0.17
Mid-size sedan	1157	0.25	0.03	0.09	0.34
Minivan/van	1157	0.07	0.02	0.02	0.13
Pickup truck	1157	0.12	0.04	0.05	0.33
Station wagon	1157	0.03	0.01	0	0.14
SUV	1157	0.11	0.02	0	0.21
Sportscar	1157	0.06	0.01	0	0.13
Premium	1157	0.09	0.03	0	0.20

8.2. Spatial regression

Table 6 displays the results for the disaggregate vehicle type choice models. Each column represents the logit model for the specified dependent variable. In Table 7, the models for vehicles with statistically significant spatial lags are shown alongside their conventional counterparts. Standard errors are reported in parentheses.

In discrete choice models, but also for nonlinear models more generally, parameter estimates do not themselves signify the magnitude of the relationship between the explanatory variables and the outcome. Instead, the coefficients in the table indicate the sign and significance of the probability of selecting the applicable vehicle type, given an increase in the regressor. For example, the coefficient for household size in Table 6, column 2 is −0.27. As it is highly significant, each additional household member lowers the likelihood of a given household purchasing a compact car, compared to other body types. One candidate substitute are minivans, which exhibit a similarly significant coefficient, except positively signed.

The results presented in Table 6 are generally as expected and match well to other, similar studies (Bhat and Sen, 2006; Choo and Mokhtarian, 2004; Potoglou, 2008). Higher local population density increases the likelihood of compact vehicle purchase, while having the opposite effect on mid-size sedans and station wagons. Income is strongly correlated with the proclivity for sportscars and premium automobiles, and inversely correlated with compacts and pickup trucks, all else

Table 6
Disaggregate spatial logit model results for 5mi threshold (dependent variable: binomial outcome for auto type).

	Compact	Coupe	Lg sedan	Md sedan	Minivan	Pickup	Statnwgn	SUV	Sportscar	Premium
Weighted ownership$_j$	3.07	−0.43	10.68	4.68**	6.19	5.83**	−33.5	3.27	0.46	16.7***
	(2.59)	(9.01)	(8.32)	(2.26)	(7.33)	(2.54)	(25.21)	(4.1)	(12.42)	(3.55)
Pop density (people/sq mi)[a]	2.52***	−1.55	−0.97	−2.02**	−0.88	1.2	−10.91***	−0.61	−3.15	−0.61
	(0.86)	(1.75)	(3.11)	(1.01)	(0.9)	(0.92)	(3.62)	(1.08)	(2.78)	(3.57)
Median yr house built[a]	0.03	−0.04	0.13	−0.02	0.02	−0.1	−0.44*	0.06	−0.1	0.08
	(0.06)	(0.08)	(0.11)	(0.05)	(0.08)	(0.09)	(0.23)	(0.06)	(0.11)	(0.09)
1999 Log income	−0.53***	0.44*	0.62*	0.22	0.21	−0.41**	0.14	0.17	1.02**	1.77***
	(0.16)	(0.25)	(0.35)	(0.15)	(0.25)	(0.18)	(0.42)	(0.19)	(0.48)	(0.28)
Household vehicles	−0.09	0.42***	0.5***	−0.3***	0.13	0.31**	−1.49***	−0.16	0.64***	0.01
	(0.14)	(0.13)	(0.18)	(0.11)	(0.15)	(0.14)	(0.32)	(0.15)	(0.17)	(0.13)
Household size (people)	−0.27***	−0.26*	−0.22	0	0.67***	−0.07	0.12	−0.12	−0.2	−0.13
	(0.09)	(0.14)	(0.16)	(0.07)	(0.09)	(0.13)	(0.22)	(0.1)	(0.21)	(0.11)
Household owned	−0.14	−0.5*	−0.32	−0.14	0.52*	0.17	0.61	0.47*	−0.25	0.34
	(0.19)	(0.26)	(0.41)	(0.17)	(0.31)	(0.28)	(0.76)	(0.24)	(0.45)	(0.26)
Licensed drivers	0.34*	−0.61**	−0.13	0.23	−0.43*	−0.31	1.44***	0.27	−1.11***	−0.15
	(0.19)	(0.31)	(0.34)	(0.18)	(0.23)	(0.28)	(0.48)	(0.25)	(0.41)	(0.22)
Householder female	0.16	0.39*	0.12	0.01	−0.37*	−0.33	0.36	−0.04	0.52	0.1
	(0.16)	(0.22)	(0.28)	(0.14)	(0.21)	(0.21)	(0.44)	(0.18)	(0.33)	(0.17)
Householder age	−0.12*	−0.2**	0.47***	0.26***	−0.08	−0.13*	0.06	−0.25***	−0.03	0.27***
	(0.07)	(0.09)	(0.15)	(0.06)	(0.09)	(0.08)	(0.15)	(0.08)	(0.13)	(0.08)
Latino householder	−0.03	0	0.83	0.33	−0.37	0.12	0.42	0.17	0	−0.55
	(0.35)	(0)	(0.65)	(0.35)	(0.66)	(0.51)	(1.07)	(0.43)	(0)	(0.65)
Black householder	−0.15	0.65	0	0.46	0.02	0	0	0.58	1.42*	1.83***
	(0.58)	(0.65)	(0)	(0.5)	(0.93)	(0)	(0)	(0.59)	(0.74)	(0.6)
Asian householder	−0.31	−0.62	0.08	0.79***	0.29	−1.04**	−0.41	−0.4	−0.14	0.29
	(0.25)	(0.41)	(0.51)	(0.2)	(0.29)	(0.46)	(0.91)	(0.31)	(0.56)	(0.26)
Job: agriculture/forestry	0.34	0	1.32	0	−0.57	1.82***	0	0.09	0	−0.58
	(0.65)	(0)	(0.89)	(0)	(1.02)	(0.6)	(0)	(0.69)	(0)	(0.78)
Job: construction	−1.1**	−0.74	−0.03	−0.21	0.01	1.5***	0	0	−0.35	−0.25
	(0.52)	(0.6)	(0.64)	(0.32)	(0.39)	(0.31)	(0)	(0.39)	(0.72)	(0.42)
Job: HVAC/service	−0.28	0	0	0.6	1.32	0.59	0	0	0	0.2
	(0.9)	(0)	(0)	(0.9)	(1.2)	(1.16)	(0)	(0)	(0)	(1.45)
Job: transportation	−1.04**	−0.49	−0.32	0.53*	0.19	0.14	0	0.39	0.01	0
	(0.44)	(0.54)	(0.68)	(0.28)	(0.45)	(0.46)	(0)	(0.35)	(0.65)	(0.38)
Job: legal	−0.8	0.89	0	1.02**	−0.32	0	1.02	−0.57	0.2	0.35
	(0.71)	(0.57)	(0)	(0.42)	(0.69)	(0)	(0.98)	(0.75)	(1.09)	(0.5)
Job: finance/insurance/R.E.	0.13	0.1	0	0.12	−1.12**	−0.04	−0.07	0.33	0.09	−0.16
	(0.25)	(0.35)	(0.43)	(0.22)	(0.45)	(0.37)	(0.83)	(0.26)	(0.49)	(0.28)
Job: auto service	−0.15	−0.93	−0.56	−0.12	0.56	0.35	0	0.35	0	−0.41
	(0.45)	(1.07)	(1.19)	(0.43)	(0.59)	(0.61)	(0)	(0.48)	(0)	(0.7)
Constant	−1.26	3.38	−39.14*	−2.39	−9.53	22.49	80.87*	−15.07	6.43	−40.21**
	(12.8)	(16.46)	(22.98)	(10.35)	(14.75)	(16.97)	(42.52)	(12.69)	(23.53)	(17.47)
Observations	1157	1157	1157	1157	1157	1157	1157	1157	1157	1157

Std. errors rounded to two decimals in parentheses.
[a] For the local census block group.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 7
Comparison between spatial and conventional models for vehicles with disaggregate spatial effects (dependent variable: binomial outcome for auto type).

	Md sedan		Pickup		Premium	
	Conv.	Spatial	Conv.	Spatial	Conv.	Spatial
Weighted ownership$_j$		4.68**		5.83**		16.7***
		(2.26)		(2.54)		(3.55)
Pop density (people/sq mi)[a]	−2.08**	−2.02**	0.79	1.2	−1.82	−0.61
	(1.01)	(1.01)	(0.98)	(0.92)	(3.72)	(3.57)
Median yr house built*[a]	−0.05	−0.02	−0.03	−0.1	−0.01	0.08
	(0.05)	(0.05)	(0.08)	(0.09)	(0.08)	(0.09)
1999 Log income	0.28*	0.22	−0.51***	−0.41**	1.92***	1.77***
	(0.15)	(0.15)	(0.18)	(0.18)	(0.27)	(0.28)
Household vehicles	−0.32***	−0.30***	0.33**	0.31**	−0.02	0.01
	(0.11)	(0.11)	(0.14)	(0.14)	(0.12)	(0.13)
Household size (people)	0	0	−0.06	−0.07	−0.14	−0.13
	(0.07)	(0.07)	(0.12)	(0.13)	(0.1)	(0.11)
Household owned	−0.16	−0.14	0.22	0.17	0.28	0.34
	(0.17)	(0.17)	(0.27)	(0.28)	(0.26)	(0.26)
Licensed drivers	0.23	0.23	−0.3	−0.31	−0.14	−0.15
	(0.18)	(0.18)	(0.28)	(0.28)	(0.23)	(0.22)
Householder female	0.02	0.01	−0.31	−0.33	0.11	0.1
	(0.14)	(0.14)	(0.21)	(0.21)	(0.17)	(0.17)
Householder age	0.27***	0.26***	−0.14*	−0.13*	0.28***	0.27***
	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)	(0.08)
Latino householder	0.3	0.33	0.12	0.12	−0.71	−0.55
	(0.35)	(0.35)	(0.5)	(0.51)	(0.67)	(0.65)
Black householder	0.42	0.46	0	0	1.6***	1.83***
	(0.5)	(0.5)	(0)	(0)	(0.56)	(0.6)
Asian householder	0.82***	0.79***	−1.12**	−1.04**	0.29	0.29
	(0.2)	(0.2)	(0.45)	(0.46)	(0.26)	(0.26)
Job: agriculture/forestry	0	0	1.85***	1.82***	−0.54	−0.58
	(0)	(0)	(0.54)	(0.6)	(0.77)	(0.78)
Job: construction	−0.23	−0.21	1.56***	1.5***	−0.28	−0.25
	(0.32)	(0.32)	(0.31)	(0.31)	(0.41)	(0.42)
Job: HVAC/home service	0.55	0.6	0.54	0.59	0.13	0.2
	(0.91)	(0.9)	(1.15)	(1.16)	(1.45)	(1.45)
Job: transportation	0.55**	0.53*	0.08	0.14	0	0
	(0.28)	(0.28)	(0.46)	(0.46)	(0.37)	(0.38)
Job: legal	1.03**	1.02**	0	0	0.52	0.35
	(0.41)	(0.42)	(0)	(0)	(0.46)	(0.5)
Job: finance/insurance/R.E.	0.14	0.12	−0.05	−0.04	−0.04	−0.16
	(0.22)	(0.22)	(0.36)	(0.37)	(0.27)	(0.28)
Job: auto service	−0.12	−0.12	0.37	0.35	−0.48	−0.41
	(0.43)	(0.43)	(0.59)	(0.61)	(0.7)	(0.7)
Constant	4.86	−2.39	10.5	22.49	−22.72	−40.21**
	(9.82)	(10.35)	(15.3)	(16.97)	(16.27)	(17.47)
Observations	1157	1157	1157		1157	
AIC	1378	1376	720	716	1008	985

Std. errors rounded to two decimals in parentheses
[a] For the local census block group.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

constant. Households with more members tend to prefer minivans, and avoid compact cars. Pickup trucks, coupes, large sedans and sportscars are preferred by homes with more existing vehicles, while mid-size sedans and station wagons are more likely to represent a home's first car purchase. Younger household heads seem to prefer smaller cars like compacts and coupes, but also pickup trucks and SUVs, while older drivers select sedans and premium cars. Households with at least one occupant working in agriculture or construction are more likely to purchase a pickup truck.

We include a variable for the ownership characteristics of the local neighborhood in order to test whether household vehicle choice is spatially dependent. Even after controlling for household characteristics, models for mid-size sedans, pickup trucks, and premium cars exhibit positive and statistically significant spatial lags, meaning that households persistently cluster in purchasing such auto types. In order to verify that the lag parameter sufficiently improves the choice model for compacts and pickups, we conducted a likelihood ratio test on a vector of constraints equating the lagged model to the conventional model. In all cases, the constraints were rejected. We interpret this result to indicate that the lagged models are sufficiently different, and that they significantly improve the likelihood of observing the original data. Reinforcement is provided in Table 7, where the model selection criteria show that every spatial model improves AIC.

The spatial models offer more than an improvement to model fit. Table 7 also demonstrates that the addition of a spatial lag can change estimates for the remaining explanatory variables. Prior to accounting for spatial effects, the model for mid-size sedans finds income significant at the 6% level; afterwards, income is no longer significantly different from zero. For premium cars, the constant becomes significantly negative.

9. Conclusions

The conventional, implicit assumption of spatial randomness in vehicle choice is not validated by San Francisco Bay Area automobile data. Instead, vehicle ownership of nearly every body type is characterized by spatial autocorrelation. In addition to including common explanatory variables, accounting for spatial dependence in automobile choice can improve statistical inference, since it increases efficiency in the presence of correlated error terms, and eliminates the bias resulting from a failure to account for mutual interdependence.

We show that Bay Area residents are more likely to choose a type of new car that is favored by their neighbors. On its own, this finding justifies the consideration of spatial factors in future transportation work. However, this result also poses an identification problem, since it is difficult to distinguish mutual dependence from homogeneity of preferences in auto choice. Self-selection into a certain region by like-minded consumers is equally likely to exhibit spatial auto clustering as is the possibility that vehicle choice is spatially interdependent. Still, this research does suggest that spatial factors must be considered in order to properly estimate vehicle choice models. Otherwise, adverse results may include inappropriate model selection and improper inference. Additionally, if consumers indeed influence one another in their vehicle choices, this could have important policy implications for agencies that regulate traffic flow and road safety.

One limitation of this article is the constraints imposed by the household survey we have used. BATS did not adequately sample every portion of the San Francisco Bay area. This uneven coverage may affect the results, but it is impossible to tell in which direction, since the data are missing. For example, BATS provided sufficient information for the aggregate analysis to study only a third of the census tracts in the Bay Area. The ownership concentrations that we measured are subject to the assumption that the missing tracts displayed a similar pattern to those sampled by BATS. At the disaggregate level, BATS did not provide enough geocoding sensitivity to plot households accurately. Instead, the finest geographical point to which a surveyed household could be associated was its census block group. Another constraint imposed by BATS is that the survey does not allow differentiation between spatial dependence in the choice process and selection bias. A better spatial sample and a fuller portrait of the household vehicle portfolio may allow a more comprehensive modeling approach to future empirical pursuit of this topic.

Acknowledgments

The authors are grateful to Professor Aaron Smith, two anonymous reviewers, and seminar participants at UC Davis and the USDA Economic Research Service for their valuable comments and suggestions.

References

Adjemian, M.K., 2009. Estimating Spatial Interdependence in Automobile Choice with Multilevel Data. Ph.D. Dissertation, Department of Agricultural and Resource Economics, University of California, Davis.

Adjemian, M.K., Williams, J., 2009. Using census aggregates to proxy for household characteristics: an application to vehicle ownership. Transportation 36 (2), 223–241.

Anselin, L., 1992. Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences. Technical Report 92-10. National Center for Geographic Information and Analysis (NCGIA).

Anselin, L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models. Agricultural Economics 27, 247–267.

Anselin, L., Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D.E. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York.

Anselin, L., Bera, A., Florax, R., Yoon, M., 1996. Simple diagnostics for spatial dependence. Regional Science and Urban Economics 26, 77–104.

Anselin, L., Griffith, D.A., 1988. Do spatial effects really matter in regression analysis? Papers, Regional Science Association 65, 11–34.

Assuncao, R.M., Reis, E.A., 1999. A new proposal to adjust Moran's *I* for population density. Statistics in Medicine 18 (16), 2147–2162.

Bhat, C., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transportation Research Part B: Methodological 32 (1), 61–75.

Bhat, C.R., Sen, S., 2006. Household vehicle type holdings and usage: an application of the multiple discrete–continuous extreme value (MDCEV) model. Transportation Research Part B: Methodological 40 (40), 35–53.

Choo, S., Mokhtarian, P.L., 2004. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. Marketing Science 12, 194–208.

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

De Jong, G.C., Fox, J., Daly, A., Pieters, M., Smit, R., 2004. Comparison of car ownership models. Transport Reviews 24 (4), 379–408.

Dugundji, E.R., Walker, J.L., 2005. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. Transportation Research Record 1921, 70–78.

Florax, R.J.G.M., Vlist, A.V.D., 2003. Spatial econometric data analysis: moving beyond traditional models. International Regional Science Review 26 (3), 223–243.

Goetzke, F., 2008. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. Urban Studies 45 (2), 407–417.

Kawabata, M., Shen, Q., 2007. Commuting inequality between cars and public transit: the case of the San Francisco Bay Area, 1990–2000. Urban Studies 44 (9), 1759–1780.

Lesage, J.P., Polasek, W., 2005. Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows. Working Paper. Texas State University.

Lin, C.-Y.C., forthcoming. A spatial econometric approach to measuring pollution externalities: an application to ozone smog. Journal of Regional Analysis and Policy.
Mendenhall, W., Sincich, T., 1996. A Second Course in Statistics: Regression Analysis. Prentice Hall, Upper Saddle River, NJ.
Mohammadian, A., Haider, M., Kanaroglou, P.S., 2005. Incorporating spatial dependencies in random parameter discrete choice models. In: 84th Annual Meeting of the Transportation Research Board. Washington, DC.
Mohammadian, A., Miller, E.J., 2003. Empirical investigation of household vehicle type choice decisions. Transportation Research Record 1854, 99–106.
Nobile, A., Bhat, C.R., Pas, E.I., 1997. A random effects multinomial probit model of car ownership choice. Case Studies in Bayesian Statistics III, 419–434.
Paez, A., Scott, D.M., 2007. Social influence on travel behavior: a simulation example of the decision to telecommute. Environment and Planning A 39 (3), 647–665.
Potoglou, D., 2008. Vehicle-type choice and neighbourhood characteristics: an empirical study of Hamilton, Canada. Transportation Research Part D: Transport and Environment 13 (3), 177–186.
Robinson, W.S., 1950. Ecological correlations and the behavior of individuals. American Sociological Review 15, 351–357.
Scott, D.M., Kanaroglou, P.S., 2002. An activity-episode generation model that captures interactions between household heads: development and empirical analysis. Transportation Research Part B – Methodological 36 (10), 875–896.
US Census Bureau, 1994. Geographic Areas Reference Manual. US Bureau of the Census, Washington, DC.
Vovsha, P., Petersen, E., Donnelly, R., 2004. Impact of intrahousehold interactions of individual daily activity-travel patterns. Travel Demand and Land Use 1898, 87–97.
Zhao, Y., Kockelman, K.M., 2002. Household vehicle ownership by vehicle type: application of a multivariate negative binomial model. In: 81st Annual Meeting of the Transportation Research Board. Washington, DC.