13

40599

2009

Estimating Spatial Interdependence in Automobile Type Choice With Survey Data

Michael Adjemian, C.-Y. Cynthia Lin, Jeffrey Williams

Abstract

Using San Francisco Bay Area Travel Survey (BATS) data, this paper uses spatial econometrics to evaluate whether consumer interaction influences automobile choices. We demonstrate how to determine if space is a factor, establish whether it is true or spurious, and modify choice models in order to control for spatial effects. We provide evidence for aggregate level concentrations in the proportionate ownership of several different auto types after controlling for potential confounders. At the disaggregate level, we apply a spatially autoregressive logit model to the decision to buy a new car type. According to our results, including spatial factors can improve vehicle choice models.

Keywords Vehicle choice modeling; Spatial Econometrics; Survey data; Discrete choice
JEL Classification Codes R41, R48, C2

DRAFT: January 22, 2009
Word Count: 7462

Automobiles 168

Michael K. Adjemian[1]
Department of Agricultural and Resource Economics
University of California, Davis
2409 Charleston Oaks Lane, Decatur GA, 30030
Email: mkadjemian@ucdavis.edu
Phone: (404) 549-3711

C.-Y. Cynthia Lin
Department of Agricultural and Resource Economics
University of California, Davis
3104 Social Sciences & Humanities, University of California, Davis, One Shields Avenue, Davis CA 95616, USA
Email: cclin@primal.ucdavis.edu
Phone: (530) 752-0824

Jeffrey Williams
Department of Agricultural and Resource Economics
University of California, Davis
2144 Social Sciences & Humanities, University of California, Davis, One Shields Avenue, Davis CA 95616, USA
Email: williams@primal.ucdavis.edu
Phone: (530) 754-7625

[1] Corresponding author

Working paper.

3.1 Introduction

Although space is an integral component to transportation choices, no study has yet considered whether automobile purchases are spatially influenced. Nevertheless, the makeup of the surrounding automobile fleet may play a role in the household choice process. The composition of nearby ownership may serve to signal auto reliability, normalize perceptions in the case of a new body type like SUV's, or stimulate a desire for conformity.

Spatial interdependence is signified by the presence of spatial autocorrelation (Cliff and Ord, 1973). In the case of vehicle ownership, nearby observations display would more similar ownership characteristics than do distant ones. This alone does not prove that household utility is affected by the actions of their neighbors. Alternatively, households with similar preferences may self select for certain regions. Still, it does indicate that aggregate level models of vehicle ownership must take account of the spatial dimension, and provides justification for further inquiry at the household level.

Conventional choice models express household utility as a function of its own characteristics and the traits of available alternatives (Cameron and Trivedi, 2005), but do not allow for the possibility of inter-household interaction. Consequently, if indeed social interdependence affects the vehicle choice decision, the utility function must be restated to account for the observed actions of the surrounding community. Otherwise, the choice model is misspecified, resulting in biased and invalid estimates.

Excluding a significant spatial term biases results, but the problem grows in importance if other coefficients are impacted by the missing control. Unless accounted for, coefficients of the remaining covariates will be biased to the degree that they pick up the relationship between the outcome and the missing spatial term, resulting in incorrect inference. We display such a finding in our aggregate analysis.

Using 2000 San Francisco Bay Area Travel Survey (BATS) data, we consider whether households herd in their automobile type choice in the nine-county region. 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. At the disaggregate level, we apply an autoregressive choice model to evaluate whether spatial effects affect consumer vehicle choice.

3.2 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). Formally accounting for spatial effects allows more thorough evaluation of traditional choice problems, and may be crucial to understand and properly estimate the data generating process. Anselin and Griffith (1988) show that if spatial effects are ignored, incorrect inference may result.

Transportation behavior exhibits signs of spatial interdependence, sometimes termed herding or bandwagon effects. 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) comprise notable examples. 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 between individuals affects their likelihood of exhibiting herd behavior in selecting public transit to work. No research has yet considered whether households factor the composition of local ownership into their own auto choice.

In addition to the number of cars on the road, the degree of vehicle heterogeneity affects roadway congestion, accident rates, pollution levels, and petroleum consumption. Consequently, responsible public agencies use models that project vehicle fleet composition in order to meet policy objectives. Choice models (Bhat and Sen, 2006; Choo and Mokhtarian, 2004; Mohammadian and Miller, 2003) are commonly used to estimate the parameters relevant to vehicle type choice. Since the set of car types is categorical, we employ a choice framework.

Including a spatial component can complicate discrete choice modeling. For example, Goetzke's (2008) spatial lag term is assumed to be exogenous, and no spatial autocorrelation is allowed in the error term of the utility specification.

Mohammadian et, al. (2005), also make the simplifying assumption of an independent error term in their spatial logit specification of a residential choice model. We condition household utility on observed auto choices, and model spatial effects exogenously.

3.3 Data

Vehicle fleet ownership and socioeconomic/demographic data for 15,064 households were collected by the 2000 San Francisco Bay Area Travel Survey (BATS), commissioned by the Bay Area Metropolitan Transportation Commission (MTC). During the period February 2000 to 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 geocoding the home addresses of surveyed households. 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 U.S. Census Bureau.

Proportionate auto ownership for each tract was calculated by averaging over the vehicle types exhibited by its surveyed households. Explanatory variables were drawn from census information imported from the year 2000 United States Census

Summary File 3. According to the U.S. 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). Additionally, census tracts are intended to be permanent statistical subdivisions, increasing their usefulness in empirical applications. Census measures include population size, racial composition, average age, average educational attainment, marital status, and median income.

We classified the vehicles in BATS into nine vehicle types according to those used by the auto information company Edmunds.com, Inc.: coupe, compact sedan, mid-size sedan, large sedan, station wagon, sports utility vehicle (SUV), pickup truck, minivan/van, and sportscar. Additionally, we created two additional indicators: 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 Porsche, BMW or Ferrari. Therefore, we investigated the presence of spatial interdependence in eleven categories of car ownership.

Those tracts that did not display sufficient observations according to the definition of proportionate ownership were excluded. Out of the 1332 tracts surveyed by BATS, requiring that a tract have at least 20 cars or 10 surveyed homes limited the sample to 425 observations and 560 tract observations, respectively. Those cars in BATS that could not be readily identified or classified into car type were not included.

For the disaggregate analysis, the dependent variable is the binomial outcome associated with the purchase of a given model year 2000 body type. Explanatory variables were taken directly from BATS; census information for block group density and median housing age was imported from the year 2000 United States Census Summary File 3. Vehicle characteristics used in the disaggregate analysis were obtained from the Cars.com (a division of Classified Ventures, LLC) used car buying guide research feature. For each car, we obtained purchase price, type of drive wheels (front, rear or all wheel drive), engine displacement (in cubic inches), horsepower, Environmental Protection Agency (EPA) rated miles per gallon (for city and highway travel), and curb weight.

About 6% of BATS vehicles, or 1,660 out of 27,822 records, represented model year 2000 cars. Of these, 439 did not contain information essential to this study, such as self-reported household income, employment status, or age. Choosing to define a narrow neighborhood for each household, we settled on the smallest possible distance cutoff for the spatial weight matrix: a three-quarter mile radius. In order to produce meaningful estimates, we further eliminated 496 observations that did not contain at least 30 neighbors in that radius.

3.4 Aggregate Methods

A consequence of spatial autocorrelation in auto ownership data is that estimates generated by applied research can be adversely affected. Statistical inference from models that do not account for 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 be adequate, models themselves can be modified to control for the spatial dimension of the data. Since we rely on a previously conducted survey, and cannot increase the sample size or take other corrective measures, 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 auto choice exhibits spatial autocorrelation, and review models that relax the assumption of spatial independence—explained in detail by Anselin (1992).

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. In this fashion, aggregate, also termed ecological, travel behavior data is related to community level characteristics. For example California's Department of Transportation prepares the annual Motor Vehicle Stock, Travel and Fuel Forecast (MVSTAFF) using county level data including auto body type ownership rates, population, and income level. A risk of using ecological data is that the explanatory variables may be sufficiently correlated and pose a multicollinearity concern. Despite this, the data and computation requirements are often much easier to satisfy than those essential to a micro-level framework.

3.41 Detecting Spatial Autocorrelation

The presence of spatial autocorrelation signifies that a variable is spatially dependent. If the data are further positively spatially autocorrelated, this dependence is observable in the form of spatial clusters. For example, as shown in Figure 1, BATS data show that census tracts with like rates of pickup truck ownership are spatially congregated, and is particularly sparse in the census tracts that make up the city of San Francisco. In order to detect whether auto ownership is globally spatially autocorrelated, we compute the Moran's I statistic. A rejection of the null hypothesis of no spatial autocorrelation is evidence of spatial dependence, with a positive or negative relationship as indicated by its sign.

However, the significance of Moran's I does not imply that proportionate ownership is truly spatially dependent (Lin, 2008). Quite possibly, other factors might be the source of the spatial autocorrelation. For example, the dependence may vanish after explanatory variables are considered and a flexible error process is specified. In the case of Figure 1, does the low pickup truck ownership in San Francisco indicate spatial interdependence, or is it perhaps an artifact of high population density?

3.42 Dependent Variable: Proportionate Auto Ownership

Interdependence in vehicle choice may be evident in the rate of similar cars (vehicle rate), or the rate of households that own a similar car (household rate). Moreover, the dichotomous definition of the dependent variable serves to provide robustness to the study. As a result, we use two different dependent variables in each of the aggregate regression analyses. For a given area, the vehicle rate is calculated by a simple ratio of the cars of certain type located in a region, divided by the total number of cars in that area. The regional level considered in our analysis is that of census tract, since it was the lowest level of aggregation that allowed a useful number of observations. For a given body type i, the vehicle rate is defined by:

(1) $$y_i = \frac{N_{i,t}}{C_t}, \text{ for } i=1,2,\ldots,11 \text{ and } t=1,\ldots,T, \text{ where}$$

(2) $$N_{i,t} = \sum I(n_t = 1), \text{ for } i=1,2,\ldots,11, \text{ and}$$

(3) $$C_t = \sum_{i=1}^{i=9} N_{i,t}$$

Here, i=1,…,9 refers to the nine different body type classifications for the automobiles in BATS: compact sedan, SUV, etc. The other two values for i (10,11) indicate whether the vehicle in question can also be termed a premium or new car. N, the total number of cars of a certain type in census tract t, is the sum of n, the indicator for whether a given car in the tract is of type i. The total number of cars in the tract, C, is calculated by totaling only the number of cars that are classified body types 1-9 so as to avoid double counting. The same method is used to calculate the proportion of households owning a similar car (household rate), with N and C chosen to represent the tract's households that own a car of type *i*, and the total number of surveyed households, respectively. Since y is a proportion of cars or households, *T* is rationed by setting minimum levels of observations in order to ensure meaningful estimates.

3.43 Naïve Aggregate Model

Aggregate ownership models can be used to estimate vehicle type choice, if only for a limited amount of alternatives (De Jong *et al.*, 2004). For a given body type *i*, we define the naïve model so that it relates proportionate auto ownership to the characteristics of the region under consideration, without taking account of spatial effects. For example, the rate of pickup trucks in a region (or households owning a pickup truck) is regressed on area-wide characteristics likely correlated with auto type ownership. A general form of the aggregate type ownership model is given by a simple linear model:

$$y_i = x\beta + \varepsilon_i \tag{4}$$

where y is a Tx1 vector of dependent variables that represent the rate of auto type i ownership for every region t. Regional characteristics are represented by the Tx1 vector x, while the 1xT vector β indicates the relationship between the regressors and the outcome. 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, average educational attainment, marital status, and racial makeup. We incorporated population density, and average travel time to work since these variables may prove important, particularly from a spatial perspective. The Tx1 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).

However, if 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 $\mathrm{cov}(y,\varepsilon) = 0$, 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 the spatial dependence is present, statistical inference can be misleading.

3.44 Aggregate Error Model

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

(5) $$y_i = x\beta + \varepsilon_i \text{, where } \varepsilon_i = \lambda W \varepsilon_i + \xi_i$$

The weight matrix is a TxT matrix with zeros on the diagonal so that the error in a particular location cannot affect itself. The remaining values in W indicate the amount of influence each tract location is modeled to have on every other tract in the dataset, which we define in proportion to the Great Circle distance between the latitude and longitude coordinates provided by the US Census Bureau. 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, so it is spatially random.

If the spatial error model effectively explains the spatial dependence of the system, then efficient estimates of β can be confidently recovered. Lin (2008) 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). In that case, inference on the parameters is not adversely affected.

3.45 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 or error model adequately specifies the system. In order to avoid omitted variable bias, the spillover effect from one tract to another must be controlled for explicitly. The autoregressive aggregate model is then:

$$y_i = \rho W y_i + x\beta + \varepsilon_i \tag{6}$$

Again, the weight matrix defines the structure of spatial interdependence, and the expression Wy represents the spatially weighted average of nearby auto ownership. The weight matrix has zeros on the diagonals so that no tract can affect itself. Assuming the model is correctly specified, the significance of ρ indicates whether proportionate ownership is substantively spatially dependent. A positive ρ represents of positive externality of ownership, while a negative ρ signifies negative spatial autocorrelation, after controlling for the predictors contained 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 does not make OLS parameter estimates unbiased and consistent, since $\text{cov}(y,\varepsilon) \neq 0$ and the dependent variable is correlated with the error term. The model can be suitably estimated by ML. However, 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).

3.46 Choosing the Weight Matrix

For both the spatial lag and error models, the spatial weight matrix plays an important 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 among the observed tracts. If the data is characterized by contiguity, then neighbors can be determined on the basis of sharing a border. In this paper, the ownership data is drawn from BATS, and represents a spatial sample that does not fully cover the bay area. Consequently, the weight matrix is calculated using the spatial distance between tract centroids.

For each definition of the dependent variable, the weight matrix is computed by setting a threshold distance of twenty miles as the maximum allowable neighborhood, and calculating the inverse distance between locations. The average number of neighbors for the tracts that met the minimum requirements is 120 and 153, for the vehicle and household rate, respectively. In both cases, mean distance between neighboring tracts is 11.8 miles.

Although the weight matrices in this paper were calculated with a maximum distance threshold of twenty miles, we verified that the robustness of our results to multiple weight matrix specifications. Considering a range of threshold distances up to fifty miles, we found that our results were repeatedly confirmed. Additionally, we contemplated the use of a "k nearest neighbor" matrix, which would define the neighborhood by the number of observations (k) rather than by an arbitrary distance threshold. The nearest neighbor approach may be useful in situations that evaluate census tracts in both rural and urban areas, such as this study, since they constrain the number of neighbors to be the same (Anselin, 2002). However, the resulting asymmetric spatial weight matrix is not supported in Geoda (Anselin, 2003), the program we used to construct our aggregate weight matrices. Still, we evaluated the robustness of our results to asymmetric matrices by using the spdep regression package in the statistical program R (R Development Core Team, 2005), since that program supports weight matrix asymmetry, and found that our results were similar for nearest neighbor specifications.

3.47 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 is a crucial step, since it determines whether we include a spatial element, and affects that way that we interpret the spatial dependence in auto ownership: nuisance or substantive. A variety of methods have been proposed to answer this question (Anselin and Bera, 1998), and we present both of the tests applied in this paper.

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 design. Additionally, they 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 OLS estimation.

3.48 Durbin Model: Testing for "True" Spatial Dependence

If OLS is judged not optimal, a more elegant way to determine the proper specification is 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 equation (5), beginning by rearranging the error process:

$$(6) \qquad \varepsilon_i = \lambda W \varepsilon_i + \xi_i \qquad \rightarrow \qquad \varepsilon_i = (I - \lambda W)^{-1} \xi_i$$

Substituting into equation (5), and organizing terms:

$$(7) \qquad y_i = x\beta + (I - \lambda W)^{-1} \xi_i \qquad \rightarrow$$

$$(8)\qquad y_i = \lambda W y_i + x\beta - \lambda W x\beta + \xi_i$$

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. However, the coefficients of the Durbin model need not be constrained so that the right hand side product of the first and second terms equals the opposite of the third:

$$(9)\qquad y_i = \lambda W y_i + x\beta - \delta W x + \xi_i$$

In the literature, the nonlinear constraint that $\delta = \lambda\beta$ is referred to as the common factor hypothesis. If the constraint holds, the Durbin model in equation (9) collapses to the original spatial error specification, equation (5).

The null hypothesis is that the spatial dependence is adequately specified by an error model. After estimating both the error and Durbin models, the constraint is tested by means of a likelihood ratio test. Whether or not the null hypothesis is rejected depends on the increase in log likelihood. If the null hypothesis is rejected, it indicates that the error model does not suitably account for the spatial autocorrelation in the dependent variable.

There are some limitations, however. First, 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 OLS is satisfactory. 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.

3.49 Estimation Strategy

In order to determine whether proportionate ownership is spatially dependent, we perform LM tests on OLS errors for all eleven car types, and for both the vehicle rate and the household rate. If the LM tests suggest that a spatial model is appropriate, we conduct common factor tests to determine whether the dependence is "true" or "spurious". Taken together, we select the appropriate model, and estimate it in order to search for spatial effects, controlling for potential confounders.

3.5 Aggregate Results

3.51 Spatial Diagnostics

Although not sufficient in demonstrating substantive spatial dependence, evidence of autocorrelation signifies that the dependent variable is distributed differently than we would expect if its observation were truly random. After calculating auto ownership rates, we evaluate whether the data exhibit spatial autocorrelation by calculating Moran's I for each of the eleven auto classifications, and for both definitions of the dependent variable. Moran's I did not display commonly accepted significance levels for coupes (0.07 for the household rate, 0.29 for the vehicle rate), and vehicle rate large sedans (0.43) and sportscars (0.63). For both definitions of ownership, midsize sedans, pickup trucks, station wagons, and new cars were all positively spatially autocorrelated at a p-value lower than 0.001. The null hypothesis of no positive spatial autocorrelation was rejected at the 5% level for every other body type. These results indicate the presence of clustering in most kinds of automobile ownership in the aggregate data.

3.52 Multicollinearity Assessment

In some cases, aggregate variables are highly correlated. When these are used together as regressors, the resulting multicollinearity can confuse the sign and significance of parameter estimates. In that case, although the model is still valid, inference about individual predictors may be adversely affected. We use the variance inflation factor (VIF) method to determine how likely multicollinearity is to affect the estimates. Mendenhall and Sincich (1996) propose a cutoff rule of 10, above which multicollinearity is suspected. We calculated the VIF for the aggregate variables in our analysis and verified that they did not exceed this cutoff.

3.53 Descriptive Statistics

Table 1 displays means, standard deviations, minimum and maximum values for the variables we used to model aggregate vehicle type ownership with a household rate dependent variable. Table 2 shows the descriptive statistics for the case in case of vehicle rates. In the top panel of each table, ownership rates are provided for the different vehicle types considered; explanatory variables are listed in the bottom panel. The difference between the two tables represents the distinction in average community attributes across dependent variables, and is a result of the sampling methodology used by BATS combined with the minimum cutoff values. The first nine auto categories shown in each table are exclusive, in that each stands on its own as a separate body type. Of these nine, those vehicles that can be further classified as "premium" or "new cars" are included in the final two categories.

Ownership levels for the household rate do not sum to one, since this rate is calculated as the percentage of homes that own a similar car. On the other hand, the ownership levels for the first nine categories in the vehicle rate do sum to one, by construction. Compact and midsize sedans are the most popular vehicle types represented in BATS, owned by 35% and 40% of households, 21% and 25% of vehicles, respectively. The least popular car types across dependent variables are station wagons, sportscars and coupes. Explanatory variables are drawn from U.S. Census tract data. Between Tables 1 and 2, the largest discrepancy in the average community characteristics is that household rate census tracts have a higher population density. Additionally, they are slightly less wealthy, younger, and have a smaller household size.

3.54 Spatial Regression

Nearly every auto type in BATS is spatially autocorrelated, but this result could be interpreted incorrectly without controlling for potential confounders. We use spatial regression to evaluate whether aggregate auto ownership of any body type is spatially concentrated after including the explanatory variables in Tables 1 and 2. All regressions are estimated via maximum likelihood.

The LM decision rule indicates that OLS regressions of the household rate dependent variable on the predictors are missing a spatial lag for the pickup truck, station wagon, and SUV body types, as well as for the "new car" designation. When the vehicle rate serves as the outcome, only pickup trucks and station wagons are judged to be best represented by a spatial lag model. In every case, we verified that the spatial lag specification improved on the error format by additionally conducting a test of the common factor hypothesis, and rejected the constraints implied by the null hypothesis at the 5% significance level. Together, the LM and common factor tests indicate that for these body types, the OLS model suffers from omitted variable bias. The results for the household rate are reported in Tables 3, with the lag model parameters displayed alongside their OLS counterparts. Vehicle rate coefficients are shown in Table 4. In both tables, *t* statistics in parentheses signify the degree of statistical significance for each parameter estimate.

Since each dependent variable is continuous, the regression results can be interpreted as the percentage point increase in the rate of homes that own a given vehicle type

(Table 3) or rate of similar vehicles (Table 4) correlated with a one unit—or otherwise specified—increase in the regressor. For example, column 3 in Table 3 reports that a one hour increase in average travel time to work is associated with a 23 point increase in the proportion of homes that own a pickup truck.

The lone exception is the coefficient on the spatial lag. In order to maintain a symmetric weight matrix, which simplifies estimation, we avoided row standardization. Unfortunately, this also complicates the interpretation of the spatial coefficient. Essentially, the spatial effects in each table represent the average correlation of the weighted average of local ownership, calculated using the inverse distance from the observed census tract, with the dependent variable. Therefore, although a straightforward interpretation for a given body type is not possible, a significant coefficient on the spatial lag is a sign that auto ownership for that body type remains spatially correlated after controlling for a range of explanatory variables. In this case, since the spatial lags are all positive and significant, this means that vehicle type ownership is concentrated spatially; for these body types, nearby census tracts are more likely to exhibit common ownership characteristics than are distant ones.

Compared to traditional OLS, the spatial approach to modeling aggregate vehicle type choice offers two important advantages. First, and most importantly, since the OLS approach produces inconsistent estimates, it cannot be relied upon for correct inference. Evidence of a problem with incorrect estimates is apparent in Tables 3 and

4. Obviously, the OLS models are missing the positive and significant lagged parameter for each body type. Additionally, the OLS model parameters are biased since they do not consider the spatial dimension of the data. In Table 3, after spatial concentration is accounted for, the coefficient on logged median census tract income becomes insignificant. Also, the rate of Latino residents does not in fact correlate with stationwagon ownership, and tracts with older homes tend to be associated with higher pickup truck ownership. Although the signs and significance of the remaining coefficients are unchanged, the degree of their difference can have an impact if they serve as inputs to policy sensitive models, like Caltrans' MVSTAFF.

Another advantage offered by the spatial models is that they offer a better data fit than do the conventional models. Since it provides an additional free parameter, McFadden's pseudo R-squared value for each lagged model exceeds its OLS counterpart. The corollary is that the spatial lag models produce a likelihood gain. For each body type, 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, lag models produced a lower AIC, and thus represent the preferred method to model vehicle type ownership. Therefore, for the body types represented in Tables 3 and 4, spatially lagged aggregate ownership models produced unbiased coefficients, and better data fit when contrasted with the traditional approach.

3.6 The Ecological Fallacy

Aggregate level results are useful for agencies and decision makers concerned with the collective behavior of consumers. For example, a model that estimates vehicle choice using explanatory variables drawn from local U.S. census tracts can adequately predict choices at the tract level. However, Robinson (1950) showed that ecological relationships do not necessarily translate to the individual level. In his example, although state level literacy rates were positively correlated with the proportion of immigrants, this was due to the fact that immigrants simply elected to settle in states with high levels of literacy.

In the current application, unless homogeneity constraints are imposed, it would be premature to draw inference on household level auto choice from aggregate data. Still, the existence of spatially clustered ownership among census tracts provides some indication that spatial effects are important, and leads us to inquire about their inclusion in conventional choice models. Moreover, since disaggregate level models are more appropriate for estimating the components of the choice process, we also perform an analysis of household level spatial effects. We use BATS survey data to empirically test whether spatial influence exists at the household level.

3.7 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 vehicle attributes, household characteristics and ownership choice. 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). As such, 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 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 car-type choice is made among a known set of possibilities, econometric methods used for parameter estimation are almost exclusively discrete choice, latent variable models.

3.71 Household Spatial Lag Model

Perhaps the simplest way to conceive of auto type choice at the disaggregate level is to estimate a series of binary choice models, one for each vehicle type. Although this method may not satisfy a comprehensive approach, it can tell us what explanatory variables correlate with an increased likelihood that a given household will select a certain auto body type, and provides some indication as to whether it is influenced by network effects. 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, the 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 cars is perfectly elastic. The conditional logit model applies random utility theory to the type ownership decision and has the advantage of widespread usage in transportation applications (Ben-Akiva and Lerman, 1985; McFadden, 1974).

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

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{10}$$

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

(11) $$V_{ij} = \beta_h' x_{ih} + \beta_j' x_j$$

where x_{ih} represents the relevant characteristics of household *i*, such as income, householder age, local population density, and other demographic variables. Vehicle attributes like fuel economy, manufacturer suggested retail price (MSRP), and engine size are represented by x_j. Coefficients on those 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 spatially autoregressive term

(12) $$V_{ij} = \beta_h' x_{ih} + \beta_j' x_j + \rho' Wf\left(V_{ij}\right)$$

Here, *W* is the spatial weight matrix that defines the neighborhood for every household *i*. Since observations are assumed to not affect themselves, the weight matrix is composed of zero values on the diagonal. Thus, the weighted average of nearby vehicle type choices is represented by $Wf\left(V_i\right)$. We calculate the weighted average by defining a three-quarter mile neighborhood threshold distance, weighting each neighborhood observation equally, and determining the rate of like car ownership. Records that did not consider an adequate number of neighborhood observations were removed from consideration. Any spatial effect is then translated through the parameter ρ.

The household chooses a car body type—the dependent variable (y)—in the choice set in order to maximize utility. In the binomial case, 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

(13)

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

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

(14)

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

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 equation (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. In order to avoid this obstacle, Goetzke (2008) makes the assumption that the spatial effects in public transport decisions are exogenously determined.

We circumvent the endogeneity problem for two reasons: the nature of car purchases, and the temporal indicators in the BATS data. Unlike the decision to access mass transit, which can be changed daily, once made, the choice of which car to buy is generally fixed for a period of years. As a result, spatial spillovers in auto choices are necessarily unidirectional. Fortunately, 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.

3.8 Disaggregate Results

3.81 Descriptive Statistics

The mean, standard deviation and range of values for explanatory variables used in the disaggregate model are shown in Table 5. In the table, the area wide characteristics display a considerable scope. For example, block group population density varies from as low as 211 to nearly 200,000 people per square mile. Neighborhood age also differs substantially; the median year of housing construction spans almost 80 years.

As displayed, auto ownership averages about 2 cars, while 71% of the sample records represented home owners. Although the average home contains between 2 and 3 members, less than two are licensed to drive. Nearly half the householders are female. Minorities make up less than one in five of surveyed homes. Again, the range of values displays considerable variation. The minimum of car ownership and household members is one, by model and survey construction, respectively. However, one surveyed home owned eight automobiles, while multiple homes contained seven members. The mean householder age of 44 is bracketed by a minimum age of 19, and a maximum of 88.

3.82 Spatial Regression

Table 6 displays the results for the spatially lagged binomial disaggregate vehicle type choice models. Each column represents the logit model for the dependent variable specified. The spatially lagged models are shown alongside their conventional, or "base" case, counterparts. In order to denote statistical significance, *t* statistics are reported in parentheses below the coefficient for each parameter estimate. Those variables that did not alter the probability of choice selection were dropped and do not have coefficients in the table; the number of observations represents the total records out that were not completely determined by at least one of the variables.

In discrete choice models, but also for nonlinear models more generally, parameter estimates do not explicitly signify the extent to which independent variables influence the outcome. Instead, the coefficients in the table indicate the sign and significance of the probability of selecting the relevant outcome, given an increase in the regressor. For example, according to column 3, the significant coefficient for the female householder indicator is -0.62. According to the model, female householders are less likely to purchase a compact car, versus another body type.

We include a variable for the ownership characteristics of the local neighborhood in order to test whether household vehicle choice is spatially dependent. The spatial logit models show that, even after controlling for the explanatory variables in Table 5 and vehicle characteristics, compact sedans and pickup trucks exhibit positive and statistically significant spatial lag parameters at the 5% and 10% level, respectively. 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 both cases, the constraints were rejected. We interpret this result to indicate that the lagged models are sufficiently different, and that the base case is missing a lagged parameter.

Some of the notable elements of the tables include the coefficients for population density, log income, engine size. Although small, the positive value for the population density coefficients for compact vehicles indicates that residents of more urban areas prefer those car types. The negative income coefficient in the context of pickup truck ownership signifies that high income individuals prefer other car types. In the compact lag model, the negative parameter value for engine size indicates that all else being equal, the average consumer prefers a compact car with a smaller engine. This seems intuitive. The opposite is true for pickup trucks.

In addition to the fact that the lagged models are validated by the likelihood ratio test, they are also preferred in terms of data fit. As in any model, the likelihood of observing the original outcomes is improved by the addition of a variable, such as the lagged term. Consequently, pseudo R-squared is increased by each lag model. On the other hand, in the interest of simplicity, AIC penalizes a model that includes variables without a sufficient likelihood gain. As shown in the table, AIC is improved by both spatial lag models, signifying that the spatial lag adequately improves data fit.

3.9 Conclusions

According to our results, spatial factors affect vehicle type choice. However, this result does not serve as proof that household decisions are truly influenced by their neighbors. Instead, selection bias may steer individuals with similar preferences congregate spatially. Still, this research does suggest that spatial factors must be accounted for in order to properly estimate vehicle choice models. Those agencies that use vehicle choice models as inputs or end results would benefit by considering spatial factors. If significant spatial effects are not included, adverse results include improper inference, inappropriate model selection, and suboptimal prediction.

If ownership choices are indeed influenced by the vehicle population in the surrounding community, one reason could be that individuals consume autos conspicuously. In the aggregate analysis, we found that the ownership of pickup trucks, SUVs, and new cars are spatially concentrated in the San Francisco Bay area. Each of these car types has attractive features that could conceivably influence the choices of other decision makers.

One limitation of this paper is the constraints imposed by the household survey: restricted sample size and spatial characteristics. BATS did not adequately sample every portion San Francisco Bay area, and did not collect many observations where it did have a presence. This uneven sampling may affect the results, but it is impossible to tell in which direction, since the data is 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. Additionally, this prevented us from using a contiguity matrix to test spatial characteristics. 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. Moreover, almost a third of the households that purchased a new car were missing information vital to our MNL model.

Another constraint we faced is that BATS did not allow differentiation between spatial dependence in the choice process and selection bias. A better spatial sample, and a fuller picture of the household vehicle portfolio, would add to the validity of the results. One candidate is the California Department of Motor Vehicles (DMV) registry, which contains the set of registered vehicles for every registered driver in the state, in addition to geographical indicators. If multiple DMV snapshots can be obtained, then we can more accurately segregate households by their characteristics. In this way, changes in household ownership can be modeled. In future research, we intend to use the census proxy method with local DMV records, as proposed by Adjemian and Williams (Forthcoming), to investigate the possibility that car purchases in California are similarly influenced by spatial factors.

3.R REFERENCES

Adjemian, M.K., and Williams, J., Forthcoming. Using Census Aggregates to proxy for Household Characteristics: An Application to Vehicle Ownership, Transportation.

Anselin, L., and Griffith, D.A., 1988. Do Spatial Effects Really Matter in Regression Analysis?, Papers, Regional Science Association, 11-34.

Anselin, L., 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), 1992.

Anselin, L., Bera, A., Florax, R., and Yoon, M., 1996. Simple Diagnostics for Spatial Dependence, Regional Science and Urban Economics, 77-104.

Anselin, L., and Bera, A., 1998. Spatial Dependence in Linear Regression Models with An Introduction to Spatial Econometrics, in: A. Ullah, and D.E. Giles, (Eds.), Handbook of Applied Economic Statistics, Marcel Dekker, New York, pp. 237-289.

Anselin, L., 2002. Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models, Agricultural Economics, 247-267.

Anselin, L., Geoda 0.9 User's Guide, Spatial Analysis Laboratory, University of Illinois, Champaign, IL, 2003.

Ben-Akiva, M., and Lerman, S., 1985. Discrete Choice Analysis: Theory and APplication to Travel Demand. MIT Press, Cambridge, MA.

Bhat, C., and Pulugurta, V., 1998. A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions, Transportation Research Part B: Methodological 32, 61-75.

Bhat, C., and 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, 35-53.

Cameron, A.C., and Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Choo, S., and 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., and Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

De Jong, G.C., Fox, J., Daly, A., Pieters, M., and Smit, R., 2004. Comparison of Car Ownership Models, Transport Reviews 24, 379-408.

Dugundji, E.R., and 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., and Vlist, A.v.d., 2003. Spatial Econometric Data Analysis: Moving Beyond Traditional Models, International Regional Science Review 26, 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, 407-417.

Kawabata, M., and Shen, Q., 2007. Commuting Inequality Between Cars and Public Transit: the Case of the San Francisco Bay Area, 1990-2000, Urban Studies 44, 1759-1780.

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

Lin, C.-Y.C., A Spatial Econometric Approach to Measuring Pollution Externalities: An Application to Ozone Smog, Working Paper, University of California, Davis, Department of Agricultural and Resource Economics, 2008.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior, in: P. Zarembka, (Ed.), Frontiers in Econometrics, Academic Press, New York, pp. 105-142.

Mendenhall, W., and Sincich, T., 1996. A Second Course in Statistics: Regression Analysis. Prentice Hall, Upper Saddle River, NJ.

Mohammadian, A., and Miller, E.J., 2003. Empirical Investigation of Household Vehicle Type Choice Decisions, Transportation Research Record 1854, 99-106.

Mohammadian, A., Haider, M., and Kanaroglou, P.S., Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models, 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

Nobile, A., Bhat, C.R., and Pas, E.I., 1997. A Random Effects Multinomial Probit Model of Car Ownership Choice, Case Studies in Bayesian Statistics III, 419-434.

Paez, A., and Scott, D.M., 2007. Social Influence on Travel Behavior: A Simulation Example of the Decision to Telecommute, Environment and Planning A 39, 647-665.

R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2005.

Robinson, W.S., 1950. Ecological Correlations and the Behavior of Individuals, American Sociological Review 15, 351-357.

Scott, D.M., and 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, 875-896.

US-Census-Bureau, Geographic Areas Reference Manual, US Bureau of the Census, Washington, D.C., 1994.

Vovsha, P., Petersen, E., and Donnelly, R., 2004. Impact of Intrahousehold Interactions of Individual Daily Activity-Travel Patterns, Travel Demand and Land Use, 87-97.

Zhao, Y., and Kockelman, K.M., Household Vehicle Ownership by Vehicle Type: Application of a Multivariate Negative Binomial Model, 81st Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.

Table 1: Descriptive Statistics for Variables used in Aggregate Spatial Model (Household Rate)

Variable	Obs	Mean	Std. Dev	Min	Max
Ownership Rates					
Compact	560	0.35	0.13	0.00	0.74
Coupe	560	0.10	0.08	0.00	0.38
Large Sedan	560	0.11	0.09	0.00	0.55
Midsize Sedan	560	0.40	0.14	0.07	0.92
Minivan/Van	560	0.12	0.09	0.00	0.50
Pickup Truck	560	0.19	0.13	0.00	0.70
Stationwagon	560	0.05	0.07	0.00	0.40
SUV	560	0.19	0.12	0.00	0.82
Sportscar	560	0.10	0.08	0.00	0.43
Premium	560	0.17	0.12	0.00	0.64
New Car	560	0.12	0.09	0.00	0.43
Explanatory Variables					
Population Density (people/sq. mi)	560	6478	6371	11	42538
Log of Median Income	560	11.19	0.32	9.87	12.04
Avg. Household Size	560	2.53	0.43	1.13	4.20
Avg. Age	560	42.79	4.91	22.48	54.12
Proportion w/ Bachelor Degree	560	0.20	0.07	0.04	0.43
Avg. Travel Time to Work (minutes)	560	29.40	4.96	13.87	49.93
Median Year Housing was Built	560	1966	13	1939	1995
Proportion of Black Residents	560	0.03	0.06	0.00	0.53
Proportion of Asian Residents	560	0.15	0.13	0.00	0.70
Proportion of Latino Residents	560	0.12	0.09	0.01	0.55
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

Table 2: Descriptive Statistics for Variables used in Aggregate Spatial Model (Vehicle Rate)

Variable	Obs	Mean	Std. Dev	Min	Max
Ownership Rates					
Compact	425	0.21	0.09	0	0.62
Coupe	425	0.06	0.04	0	0.20
Large Sedan	425	0.07	0.05	0	0.25
Midsize Sedan	425	0.25	0.09	0.03	0.56
Minivan/Van	425	0.08	0.05	0	0.32
Pickup Truck	425	0.12	0.08	0	0.41
Stationwagon	425	0.03	0.03	0	0.17
SUV	425	0.12	0.06	0	0.38
Sportscar	425	0.06	0.05	0	0.21
Premium	425	0.11	0.07	0	0.35
New Car	425	0.07	0.05	0	0.25
Explanatory Variables					
Population Density (people/sq. mi)	425	4839	4035	11	32802
Log of Median Income	425	11.24	0.31	10.37	12.04
Avg. Household Size	425	2.58	0.40	1.13	4.04
Avg. Age	425	43.93	4.21	25.91	54.12
Proportion w/ Bachelor Degree	425	0.20	0.06	0.05	0.38
Avg. Travel Time to Work (minutes)	425	29.32	5.10	13.87	49.93
Median Year Housing was Built	425	1967	12	1939	1995
Proportion of Black Residents	425	0.03	0.05	0	0.53
Proportion of Asian Residents	425	0.14	0.12	0	0.70
Proportion of Latino Residents	425	0.12	0.09	0.01	0.55
Proportion of Female Residents	425	0.51	0.02	0.41	0.67
Proportion of Married Residents	425	0.48	0.06	0.19	0.62

Table3

Table 3: Aggregate Spatial Lag Model Estimation Results
(Household Rate)

	pickup		statnwgn		suv		newcar	
	OLS	Lag	OLS	Lag	OLS	Lag	OLS	Lag
Weighted Ownership		0.57***		0.42**		0.34*		0.47***
		(5.51)		(2.14)		(1.79)		(2.85)
Pop Density (+100000/sq mi)	-0.31***	-0.34***	-0.11	-0.1	-0.01	-0.02	-0.02	-0.03
	(2.64)	(2.99)	(1.59)	(1.47)	(0.06)	(0.17)	(0.23)	(0.3)
Log Median Income	-0.08***	-0.04	-0.03*	-0.02	0	-0.01	0.05**	0.03
	(2.58)	(1.12)	(1.78)	(1.29)	(0.04)	(0.32)	(1.96)	(1.02)
Avg. HH Size (people)	0.06***	0.06***	0.02	0.01	0.08***	0.08***	0.01	0.01
	(2.96)	(2.81)	(1.34)	(1.2)	(3.88)	(3.93)	(0.39)	(0.64)
Avg. Age (+10 yrs)	-0.02	-0.03	0.01	0.01	0.02	0.02	-0.01	-0.01
	(0.26)	(0.26)	(0.39)	(0.39)	(0.74)	(0.74)	(-1.65)	(-1.65)
Pct. Bachelor Degrees	-0.5***	-0.39***	0.12	0.1	0.47***	0.5***	-0.01	0.03
	(3.44)	(2.82)	(1.37)	(1.22)	(3.19)	(3.42)	(0.1)	(0.29)
Avg. Travel Time (+1hr)	0.14**	0.23***	-0.11***	-0.1***	0	0	-0.05	-0.02
	(2.31)	(3.83)	(3.12)	(2.84)	(0.04)	(0.06)	(0.98)	(0.32)
Median Yr House Built (+10yr)	-0.01	-0.02***	-0.01***	-0.01***	0.01**	0.01**	0.01***	0.01**
	(1.33)	(3.34)	(2.65)	(2.66)	(2.49)	(2.08)	(2.85)	(2.46)
Pct. Black Residents	-0.12	-0.06	0.06	0.08	-0.17*	-0.16*	-0.02	-0.01
	(1.25)	(0.63)	(1.17)	(1.45)	(1.76)	(1.67)	(0.34)	(0.19)
Pct. Asian Residents	-0.27***	-0.19***	-0.02	-0.01	-0.16***	-0.15***	0.03	0.01
	(5.77)	(3.95)	(0.8)	(0.54)	(3.27)	(3.31)	(0.97)	(0.31)
Pct. Latino Residents	-0.01	0.05	-0.09**	-0.07	-0.03	-0.05	0.08	0.04
	(0.18)	(0.69)	(2.06)	(1.61)	(0.41)	(0.62)	(1.43)	(0.74)
Pct. Female Residents	-0.89***	-0.79***	0.14	0.13	-0.25	-0.24	0.04	0.05
	(3.89)	(3.58)	(1.01)	(1)	(1.06)	(1.04)	(0.22)	(0.3)
Pct. Married	0.18	0.14	-0.01	0	0.07	0.08	0.08	0.1
	(1.29)	(1.05)	(0.11)	(0.01)	(0.53)	(0.6)	(0.76)	(0.94)
Constant	2.67***	3.81***	1.72***	1.57***	-2.32**	-1.93**	-2.36***	-1.92***
	(2.83)	(4.14)	(3.06)	(2.83)	(2.39)	(1.98)	(3.26)	(2.67)
Observations	560	560	560	560	560	560	560	560
Pseudo R-squared	0.36	0.39	0.07	0.08	0.18	0.18	0.09	0.10
AIC	-899	-922	-1479	-1482	-866	-867	-1196	-1201

t statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Aggregate Spatial Lag Model Estimation Results (Vehicle Rate)

	pickup		statnwgn	
	OLS	Lag	OLS	Lag
Weighted Ownership		0.48***		0.34*
		(4.23)		(1.67)
Pop Density (+100000/sq mi)	-0.27**	-0.21*	-0.04	-0.02
	(2.37)	(1.93)	(0.66)	(0.41)
Log Median Income	-0.09***	-0.06**	-0.01	-0.01
	(4.14)	(2.53)	(0.88)	(0.61)
Avg. HH Size (people)	0	0	0	0
	(0.05)	(0.31)	(0.12)	(0.12)
Avg. Age (+10 yrs)	0	-0.01	0	0
	(0.26)	(0.26)	(0.39)	(0.39)
Pct. Bachelor Degrees	-0.33***	-0.27***	0.08	0.07
	(3.15)	(2.65)	(1.46)	(1.31)
Avg. Travel Time (+1hr)	0.08*	0.13***	-0.04**	-0.04**
	(1.92)	(3.17)	(2.02)	(1.97)
Median Yr House Built (+10yr)	-0.01*	-0.01***	-0.01***	-0.01***
	(1.75)	(3.06)	(4.11)	(3.84)
Pct. Black Residents	-0.05	-0.05	0.05	0.05
	(0.71)	(0.73)	(1.4)	(1.49)
Pct. Asian Residents	-0.13***	-0.09***	-0.01	-0.01
	(3.87)	(2.65)	(0.7)	(0.66)
Pct. Latino Residents	0	0.04	-0.03	-0.02
	(0.06)	(0.64)	(0.97)	(0.69)
Pct. Female Residents	-0.66***	-0.61***	0.08	0.07
	(3.99)	(3.78)	(0.94)	(0.91)
Pct. Married	0.15	0.11	0.04	0.04
	(1.56)	(1.16)	(0.84)	(0.95)
Constant	2.64***	3.03***	1.45***	1.31***
	(3.63)	(4.32)	(4)	(3.62)
Observations	425	425	425	425
Pseudo R-squared	0.38	0.40	0.13	0.14
AIC	-1105	-1118	-1697	-1698

t statistics in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Descriptive Stats For Variables Used in Disaggregate Model

Variable	Obs	Mean	Std. Dev	Min	Max
Census Blkgrp Characteristics					
Pop Density (pop/sq mi)	725	9780	10298	211	172898
Median Yr House Built	725	1964	15	1939	1998
HH Characteristics					
1999 Log Income	725	11.40	0.50	8.52	12.01
Household Vehicles	725	2.11	0.85	1	8
Household Size (people)	725	2.53	1.19	1	7
Household Owned	725	0.71	0.46	0	1
Licensed Drivers	725	1.87	0.56	0	4
Householder Female	725	0.47	0.50	0	1
Householder Age	725	44	13	19	88
Latino HH	725	0.04	0.19	0	1
Black HH	725	0.01	0.10	0	1
Asian HH	725	0.13	0.34	0	1

Table 6: Disaggregate Logit Model Estimation Results (Dependent Variable: Binomial Outcome for Auto Type)

	compact		pickup	
	Base	Lag	Base	Lag
Weighted Ownership		3.78**		11.88*
		-2		-1.65
Pop Density (people/sq mi)	0.00**	0.00*	0	0
	(2.43)	(1.79)	(-0.23)	(0.17)
MSRP/ Log Income	0.00	0.00	-0.02***	-0.02***
	(0.48)	(0.39)	(-5.47)	(-5.49)
Log Income	-0.03	0	-2.79**	-2.74**
	(-0.11)	(0.01)	(-2.50)	(-2.48)
Household Vehicles	(0.08)	(0.08)	(0.69)	(0.76)
	(-0.36)	(-0.35)	(-1.21)	(-1.28)
Household Size (people)	0.06	0.12	-0.19	-0.22
	(0.36)	(0.73)	(-0.53)	(-0.57)
Household Owned	0.19	0.28	0.55	0.71
	(0.54)	(0.78)	(0.47)	(0.58)
Licensed Drivers	0.05	0.05	0.55	0.36
	(0.14)	(0.14)	(0.55)	(0.36)
Median Yr House Built (Blkgrp)	0	0	0.02	0.02
	(-0.36)	(-0.06)	(0.70)	(0.54)
Householder Female	-0.62**	-0.62**	-1.01	-1.33
	(-2.24)	(-2.23)	(-1.25)	(-1.52)
Householder Age	-0.01	-0.01	0	-0.02
	(-1.05)	(-0.76)	(-0.06)	(-0.36)
Latino HH	0.07	0.23	0.27	0.04
	(0.11)	-0.34	(0.19)	(0.03)
Black HH	-1.82	-1.67	-	-
	(-1.19)	(-1.04)	-	-
Asian HH	-0.33	-0.26	0.51	0.48
	(-0.80)	(-0.62)	(0.25)	(0.22)
Unemployed	-0.03	-0.08	0.8	0.85
	(-0.10)	(-0.23)	(0.79)	(0.80)
All Wheel Drive	0.42	0.47	2.60**	3.04**
	(0.59)	-0.66	(2.10)	(2.35)
Engine Size (inches cubed)	-0.05***	-0.05***	0.09***	0.09***
	(-4.75)	(-4.74)	(4.18)	(4.09)
HP/ Weight	77.73***	81.06***	-170.27***	-158.54***
	(3.09)	(3.22)	(-3.08)	(-2.79)
Fuel Efficiency (mi/gal)	0.31***	0.32***	-0.41***	-0.43***
	(5.14)	(5.32)	(-3.20)	(-3.18)
Constant	0.19	-7.95	15.43	24.19
	(0.01)	(-0.38)	(0.26)	(0.40)
Observations	721	721	713	713
Pseudo R-squared	0.488	0.494	0.860	0.866
AIC	403	401	98	97

Figure 1

Figure 1: Pickup Truck Ownership in the SF Bay Area