How urban sprawl matters: a joint modelling of households' vehicle choice and vehicle usage

Hao Audrey Fang *

Abstract

The paper seeks to discover how urban sprawl, measured by residential density, influences households' vehicle fuel economy choices and usage choice. I propose a new method of modelling vehicle holdings in terms of number of vehicles in each category, with a Bayesian framework of multivariate ordinal response system, circumventing the complications inherent with the traditional nested logit models. I also combine the bi-variate ordered equations with bi-variate tobit equations to jointly estimate the discrete/continuous vehicle type/usage demand in a reduced form, to offer an improvement upon the traditional discrete/continuous analysis. The estimation results would enhance our understanding of how urban policy can be utilized to shape households' vehicle-holdings compositions towards fuel efficient ones, as well as less vehicle usage.

Keywords: urban sprawl; residential density; vehicle choice; fuel economy; multivariate ordered probit; multivariate tobit; discrete/continuous.

1 introduction

The relation between urban form and travel behavior is crucial to urban planning initiatives and incentives aimed at reducing street traffic, road damage, gasoline consumption and carbon dioxide pollution. Besides discouraging vehicle travel, these reductions can also be attained by encouraging utilization of fuel efficient vehicles. For a fixed amount of travel, fuel efficient vehicles use less gasoline and produce less carbon dioxide. In addition, fuel efficient vehicles also turn out to have smaller size and less weight, which would lessen traffic burden and road damage.

This project aims to examine the degree to which households' vehicle fuel efficiency choices can be influenced by residential density, an important measure of urban sprawl.

*I would like to thank David Brownstone, Jan Brueckner, Ivan Jeliazkov, Dale Poirier, and Kurt Van Dander for their guidance and helpful comments. All remaining errors are mine. Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100. E-mail: hfang@uci.edu.

The results will contribute to our knowledge of how urban policies could be tailored so that households tend to choose fuel efficient vehicles, as well as driving less. A well developed and high density area tends to have smaller parking spaces, narrower streets and more severe traffic. These conditions all work in favor of choosing smaller, easy to maneuver and more fuel efficient vehicles. Therefore, we would expect urban sprawl have a negative impact on choice of fuel efficient vehicles.

Due to potential policy implications it may offer to reduce congestion and fuel consumption, the relation between urban form and travel behavior has been a constant subject of research in urban and transport economics. Residential density, as an important aspect of urban form and a measurement of urban sprawl, has been intensively analyzed with its relation to mode choice, trip rates, and vehicle usage. It is found to be associated with higher level of transit use(Dunphy and Fisher 1996), lower fuel consumption and milage per vehicle(Golob and Brownstone 2005), and lower trip rates(Cervero and Kockelman 1991). This paper concentrates on the relationship between residential density and household's choice on vehicle type choice, which is motivated by the results from Golob and Brownstone(2005), and the relationship between residential density and vehicle usage. Golob and Brownstone discovered that fuel consumption per vehicle declines more sharply than mileage per vehicle with increasing housing density, indicating a positive relationship between fuel efficiency and housing density. There are several reasons why this could happen: people in low density area tend to drive hectically, peddling on the break harder; or low density areas are more likely to be mountain or muddy areas where mpg is lower compare to smooth urban roads with the same driving behavior, or simply some properties of high density areas as mentioned induce its residents to choose vehicles with better fuel economy. If the last one is true, then the discovery will shed lights on policy implications to make people choose fuel efficient vehicles.

1.1 vehicle type choice models

The traditional method of studying vehicle holdings decision exclusively focuses on two-stage sequential decision: first to decide total number of vehicles, and second to choose vehicle types given the total number. The sequential decision is usually estimated by a nested logit model(Train 1984, Berkovec and Rust 1985, Mannering and Winston 1985, Goldberg 1998, Feng, Fullerton and Gan 2005, West 2005). With the number of vehicles held by each household increases, vehicle type choice set proliferates. For example, if vehicles are classified into five categories there will be altogether 15 kinds of vehicle combination in a two-vehicle household and 35 kinds of vehicle combination in a three-vehicle household. The nested logit model becomes incapable in handling this proliferation in the size of vehicle type combination. Due to estimation complications arise from studying more than two vehicle holdings, most of these studies only concentrate on one-vehicle and two-vehicle households. This results in a loss of useful data. For example, in 2001 National Household Travel Survey(NHTS) data set, one-vehicle and two-vehicle households comprise of only 55 percent of the total households surveyed. Omitting the rest 45 percent with more than two vehicles will lead to biased estimates.

Since vehicle holdings composition is fully represented by number of each vehicle type, I propose to use number of each type of vehicles, instead of total number of all vehicles, to model households' vehicle holdings decision. Because the number of each vehicle type is ordered and the choices of each vehicle type within one household are interrelated, I utilize a multivariate ordered probit model with a correlated covariance structure. This model will add tremendous flexibility to vehicle holdings choice model.

1.2 Discrete/continuous models

In the previous literature of discrete/continuous models, researchers have tried hard to derive the discrete and continuous demands from utility maximization. A conditional indirect utility function, giving the maximum utility achievable, provides an important link for deriving continuous demand for vehicle miles and discrete demand for vehicle type. The drawbacks of this approach are: a) the treatment of the error term in the utility function that determines discrete choice is not innocuous. Many used Taylor expansion and ignored variability caused by categories. b) the estimation procedure is complicated. The indirect utility function derived from Roy's identity is in a non-linear form. Researchers either added some assumptions to get a linear equation(Dubin and McFadden 1984) or used a linear approximation(Train 1987, Goldberg 1998, West 2004). So the elegance from utility maximization derivation is dulled by the approximation in estimation. By uniting bi-variate ordered probit and tobit model and assume an unrestricted covariance matrix, we can cleanly estimate a reduced-form discrete/continuous system. The estimation is aided by data augmentation and Bayesian Markov Chain Monte Carlo methods.

1.3 Theoretical motivation on vehicle type choice

Let us assume a continuum of N identical risk-neutral consumers. To simplify matters, nature assigns these N consumers to a city with size A, so that consumers' choice of which city to reside in is uncorrelated to their choice of vehicles. Assume the utility from choosing vehicle size w is given by a continuous and concave function $u(w)$. Since vehicle size is closely related to weight and engine displacement, in addition to the seating comfort the utility comes also from safety assurance and acceleration. The cost of holding a w size vehicle comprises capital cost, operation cost, parking-searching cost, and extra cost of effort in manoeuvering the vehicle in highly-populated areas. Let $\bar{p}$ be cost per unit of vehicle size in terms of capital and operational cost. Let $s(\frac{w_i \cdot \sum_{i=1}^{N-1} w_i}{A})$ be cost of manoeuvering and searching for parking space. The cost depends not only on own vehicle size, but also on city size, total size of all the other vehicles driving on the streets and searching for parking. $s(\cdot)$ is a smoothly increasing function. In equilibrium, each consumer would choose w_i such that the marginal benefit equals marginal private cost:

$$u'(w_i) = \bar{p} + s' \cdot \frac{\sum_{i=1}^{N-1} w_i}{A} \quad (1)$$

Since consumers are identical, the equilibrium choice of w_i's are equal. The above equation can also be written as:

$$u'(w_i) = \bar{p} + (N-1) \cdot \frac{s'w_i}{A} \quad (2)$$

When the city size A increases, the marginal benefit of vehicle size decreases. With the same population N, this indicates that consumers in a less dense area would choose larger vehicles, and vise versa. The equilibrium, however, is not social optimal. In the equilibrium, a consumer does not consider the cost of an extra size of vehicle imposed on the other consumers, hence marginal private cost is lower than marginal social cost. Follow Brueckner and Largey (2006), marginal social cost is found by imposing symmetry in the cost function at the outset, $\frac{d(\bar{p}w_i + s(\frac{(N-1)w_i^2}{A}))}{dw_i} = \bar{p} + 2(N-1) \cdot \frac{s'w_i}{A}$. The social optimum is satisfied if

$$u'(w_i) = \bar{p} + 2(N-1) \cdot \frac{s'w_i}{A} \quad (3)$$

The distortion, or the difference between marginal social cost and marginal private cost, is a justification for a vehicle size tax, which is equal to:

$$t = (N-1) \cdot \frac{s'w_i}{A} \quad (4)$$

When population is large enough, the tax is approximate to:

$$t = density \cdot s'w_i^* \quad (5)$$

The distortion is larger in low density areas only when vehicle size is disproportionably larger, i.e. the product of density and size, in low density areas than high density areas. While that fact is hard to establish empirically, if we consider the hypothesis(Small and Parry 2005) that gasoline in the United States is underpriced, the distortion of gasoline consumption through use of larger size, fuel inefficient vehicles is more prominent in low density areas.

2 The Model

2.1 A multivariate ordered probit and tobit system

The data analysis is complicated in two aspects. First, vehicle choices within one household are interdependent. Instead of using univariate ordered probit model, bi-variate ordered probit model is adopted with an unrestricted covariance matrix. In addition, vehicle usage are interdependent in itself and with vehicle choices. Annual milage of cars and trucks equations are then added to the two bi-variate ordered probit equations. Second, the observation on average annual milage of cars and trucks are censored. About 56.7% of the households in our sample don't hold trucks, and 10.3% don't hold cars, with whom we observe zero milage. In a censored regression model where a large proportion of dependent variables are zero, the OLS estimates fail to account for the qualitative difference between zero observations and continuous observations, and tend to be biased(Greene, 1997). Tobit models (Amemiya 1984, Tobin 1958) has been widely used to model censored, correlated, bivariate data, and will be adopted in the analysis here.

Assume that two latent continuous variables y_1^* and y_2^* represent preference level for holding cars and trucks. Indexing household by i, $i = 1, ...N$, the system for discrete choice of number of vehicles is:

$$y_{1i}^* = \mathbf{x_i}'\beta_{11} + w_i'\beta_{12} + \epsilon_{1i} \quad (6)$$

$$y_{2i}^* = \mathbf{x_i}'\beta_{21} + w_i'\beta_{22} + \epsilon_{2i} \quad (7)$$

where x_i is a vector of characteristics of household i; The observed ordinal variables y_1 and y_2, counting number of cars and trucks held, take on values determined by the interval in which y_i* lies. $y_j = 0$, if $y_j^* \leq \alpha_1$, $y_j = 1$, if $\alpha_1 < y_j^* \leq \alpha_2$, $y_j = 2$ or more, if $y_i^* > \alpha_2$. For simplicity, let $y_j = 2$ represents holding of two or more vehicles.

For identifiability, we must either constrain the lowest and highest thresholds, α_1 and α_2, or constrain one of the threshold and the variance of ϵ_{ji}. In the latter case, we will have a constrained covariance matrix where all diagonal elements are one's, which is difficult to draw. In the former case, as suggested by Webb and Forster(2006), the constraints can be: $\alpha_1 = \Phi^{-1}(1/3)$ and $\alpha_2 = -\Phi^{-1}(1/3)$, which are same across the J equations. The equivalence between restriction of one cut point and the variance, and that of two cut points in ordered probit case is proved in Nandrum and Chen(1996) using re-parameterizations. In the case of ordered probit with three categories, only two cut-points are needed. By fixing the cut-points, parameters need estimations are only the coefficients and the covariance matrix. In this analysis, I adopt the re-parameterizations procedure to fix the two cut-point to 0 and 1. In fact, the two cut-points can be fixed to be any number, and the covariance matrix will change accordingly, hence the inference remains the same after standardization.

Next, assume that two latent variables y_3^* and y_4^*, representing uncensored and censored milage of cars and trucks respectively, can also be modelled in the same fashion:

$$y_{3i}^* = \mathbf{x_i}'\beta_{31} + w_i'\beta_{32} + \epsilon_{3i} \tag{8}$$

$$y_{4i}^* = \mathbf{x_i}'\beta_{41} + w_i'\beta_{42} + \epsilon_{4i} \tag{9}$$

Annual milage of cars y_3 is observed only when household holds at least one car; that is,

$$y_3 = y_3^*, \text{if } y_1 = 1 \text{ or } 2 \tag{10}$$

$$= 0, \text{ if } y_1 = 0 \tag{11}$$

The same logic applies to milage of trucks y_4:

$$y_4 = y_4^*, \text{if } y_2 = 1 \text{ or } 2 \tag{12}$$

$$= 0, \text{ if } y_2 = 0 \tag{13}$$

The whole system can then be written into a SUR (seemingly unrelated regression) form with four equations. The error structure is:

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix} \sim^{i.i.d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix} \right) = N(\mathbf{0}, \mathbf{\Sigma}). \tag{14}$$

The likelihood function is given as the following,

$$\begin{aligned}
& L(\beta, \Sigma; y_1, y_2, y_3, y_4) \\
\propto & \prod_{i \ni y_{1i}=0, y_{2i}=0} Pr(y_{1i}^* < \alpha_1, y_{2i}^* < \alpha_1) \\
\times & \prod_{i \ni y_{1i}=0, y_{2i}=1} Pr(y_{1i}^* < \alpha_1, \alpha_1 < y_{2i}^* < \alpha_2, y_{4i} = y_{4i}^*) \\
\times & \prod_{i \ni y_{1i}=0, y_{2i}=2} Pr(y_{1i}^* < \alpha_1, y_{2i}^* > \alpha_2, y_{4i} = y_{4i}^*) \\
\times & \prod_{i \ni y_{1i}=1, y_{2i}=0} Pr(\alpha_1 < y_{1i}^* < \alpha_2, y_{2i}^* < \alpha_1, y_{3i} = y_{3i}^*) \\
\times & \prod_{i \ni y_{1i}=1, y_{2i}=1} Pr(\alpha_1 < y_{1i}^* < \alpha_2, \alpha_1 < y_{2i}^* < \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^*) \\
\times & \prod_{i \ni y_{1i}=1, y_{2i}=2} Pr(\alpha_1 < y_{1i}^* < \alpha_2, y_{2i}^* > \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^*) \\
\times & \prod_{i \ni y_{1i}=2, y_{2i}=0} Pr(y_{1i}^* > \alpha_2, y_{2i}^* < \alpha_1, y_{3i} = y_{3i}^*) \\
\times & \prod_{i \ni y_{1i}=2, y_{2i}=1} Pr(y_{1i}^* > \alpha_2, alpha_1 < y_{2i}^* < \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^*) \\
\times & \prod_{j \ni y_{1i}=2, y_{2i}=2} Pr(y_{1i}^* > \alpha_2, y_{2i}^* > \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^*)
\end{aligned}$$

This likelihood function combined with the prior density produces a posterior density that is analytically untractable, highly dimensional and computational intensive. With data augmentation (Alber and Chib 1993, Chib and Greenberge 1998), however, y_j^* instead of y_j $j = 1, ...4$ are sampled from their full conditional distribution, the Bayesian estimation problem can be much simplified.

2.2 Posterior simulations

Given a sample of N observations on the four dependent variables $y_{ij}, i = 1, ..., N, j = 1, ..., 4$ and a prior density $f(\beta, \Sigma)$, the Gibbs sampling algorithm for a stable posterior distribution can be applied to three blocks: $\mathbf{y_i^*}|\mathbf{y_i}, \beta, \Sigma$, $\beta|\mathbf{y_i^*}, \Sigma$, and $\Sigma|\mathbf{y_i^*}, \beta$. The full conditionals for β and Σ have familiar closed forms:

$$\beta|\mathbf{y_i^*}, \Sigma \sim \mathscr{N}(\bar{\beta}, \bar{V}) \tag{15}$$

$$\Sigma|\mathbf{y_i^*}, \beta \sim \mathscr{I}\mathscr{W}(\nu + N, (\sum_{i=1}^{N}(\mathbf{y_i^*} - \mathbf{x_i}\beta)'(\mathbf{y_i^*} - \mathbf{x_i}\beta) + Q)^{-1}) \tag{16}$$

where $\bar{V} = (V_0^{-1} + \sum_{i=1}^{N} \mathbf{x_i'}\Sigma^{-1}\mathbf{x_i})^{-1}$ and $\bar{\beta} = \bar{V}(V_0^{-1}\beta_0 + \sum_{i=1}^{N} \mathbf{x_i'}\Sigma^{-1}\mathbf{y_i^*})$.

Sampling full conditional of latent variables $\mathbf{y_i^*}$ requires a bit more computation, but is equally straightforward. First, expand $f(\mathbf{y_i^*}|\mathbf{y_i}, \beta, \Sigma)$, simplied as $f(\mathbf{y_i^*}|\circ)$ in notation, according to law of total probability:

$$f(\mathbf{y_i^*}|\circ) = f(y_{1i}^*|\circ) \cdot f(y_{2i}^*|y_{1i}^*, \circ) \cdot f(y_{3i}^*|y_{1i}^*, y_{2i}^*, \circ) \cdot f(y_{4i}^*|y_{1i}^*, y_{2i}^*, y_{3i}^*, \circ) \tag{17}$$

where $[y_{1i}^*|\circ] = TN_{[\alpha_{y_i}, \alpha_{y_i+1}]}(\mu_1, \sigma_{11})$, $[y_{2i}^*|y_{1i}^*, \circ] = TN_{[\alpha_{y_i}, \alpha_{y_i+1}]}(\mu_{2|1}, \Sigma_{2|1})$, $[y_{3i}^*|y_{2i}^*, y_{1i}^*, \circ] = TN_{[-\infty,0]}(\mu_{3|1,2}, \Sigma_{3|1,2})$, $[y_{4i}^*|y_{3i}^*, y_{2i}^*, y_{1i}^*, \circ] = TN_{[-\infty,0]}(\mu_{4|1,2,3}, \Sigma_{4|1,2,3})$. The marginal and conditional means and covariance matrix of the distributions above are calculated according to Poirier (1995, P. 122).

3 Data and Sample

The study utilizes data from California sub-sample of the 2001 National Household Travel Survey (NHTS). The whole sample is composed of three files: 1) household file, 2) vehicle file, and 3) person file. Household file provides information on the demographics of each household and various measures of land use density. Vehicle file provides data on make/model and year of manufacture for every vehicle held by the households. Additional travel activity data, such as congestion condition of the highway and road, and distance to work, is obtained from the person file. The sample use in the analysis has total observation of 1569 households.

Table 1: Descriptive Statistics

Variable (observation: 1569)	Mean (Std.)
Residential density	2611.9 (1902.2)
Number of bikes	0.918 (1.180)
Total number of people	2.598 (1.404)
Number of adults	1.916 (0.741)
Total number of vehicles	1.87 (0.853)
If respondent has completed high school	31.4%
If respondent has a bachelor's degree or higher	41.6 %
If the youngest child is under 6	17 %
If the youngest child is between 6 and 15	17.5%
If the youngest child is over 21	5%
If MSA has rail	66 %
If respondent is employed	56.1%
If respondent is retired	20%
If respondent is a home-maker	10.3%
If household reside in urban area at census track level	34.2%
If household reside in a second city at census track level	18.2%
If household reside in suburban at census track level	40%
If annual household income is between 20k and 30k	11%
If annual household income is between 30k and 50k	21.4%
If annual household income is between 50k and 75k	20.1%
If annual household income is between 75k and 100k	12.6%
If annual household income is greater than 100k	21.8%
If the household owns home	69.2%

Note: Residential density is measured in housing units per square mile

Explanatory variables include density, other neighborhoods characteristics, and household demographic characteristics. Density is measured by housing units per square mile at the census block level and jobs per square mile at the track level. To capture local transit network and non-mobile facility, an indicator of whether or not MSA has rail and number of bicycles in the households are used. Demographic variables include total household annual income, respondent's education level, employment status, number of adults, number of children, children age, whether or not owning home, and zone type of the residence area.

The definitions and sample statistics of the explanatory variables used in the analysis are presented in Table 1.

Residential density is measured by housing units per square mile. The mean density in California is 2612. In the four MSAs of the California state, Los Angeles-Riverside-Orange County has the highest density with 3016 housing units per square mile on average, immediately followed by San Francisco-Oakland-San Jose with 2905 housing units. The density dropped to 2563 in San Diego and further down to 2113 in Sacramento.

Table 2: Dependent Variables

Variable	Mean	(std.)	Min	Max
number of cars held	1.3	(0.78)	0	6
number of trucks held	0.5	(0.7)	0	6
			25 quantile	75 quantile
average miles driven by cars	9436.7	(8376.1)	3138.9	10340.5
average miles driven by trucks	8307.4	(9590.315)	4357.8	12414.1

Examining the two-way table for observations at different categories, we can see that all the combination space are well-explored. In the most explored case, 475 households have one car and no truck, while the least explored case has 33 households , which hold two cars and two trucks. We can expect that the thresholds of the bi-variate ordered probit model will be identified.

Tabulate vehicle counts

number of cars	number of trucks			
	0	1	2 or more	Total
0	0	95	67	162
1	475	319	46	840
2 or more	415	119	33	567
Total	890	533	146	1569

4 Empirical results

The estimation results are reported in Table 7. The average predicted probabilities for choosing zero, one, two and more cars within a household are 0.0329, 0.6622, and 0.3049 respectively; those for trucks are 0.5258, 0.4604, and 0.0138. The covariance matrix shows a negative relation between number of cars held and number of trucks held, as well as miles driven by cars and miles driven by trucks, indicating a substitution effect between cars and trucks, not only type-wise but also usage-wise. From Figure 1, we can see that the matrix converges fast and well.

Table 3: Covariance Matrix

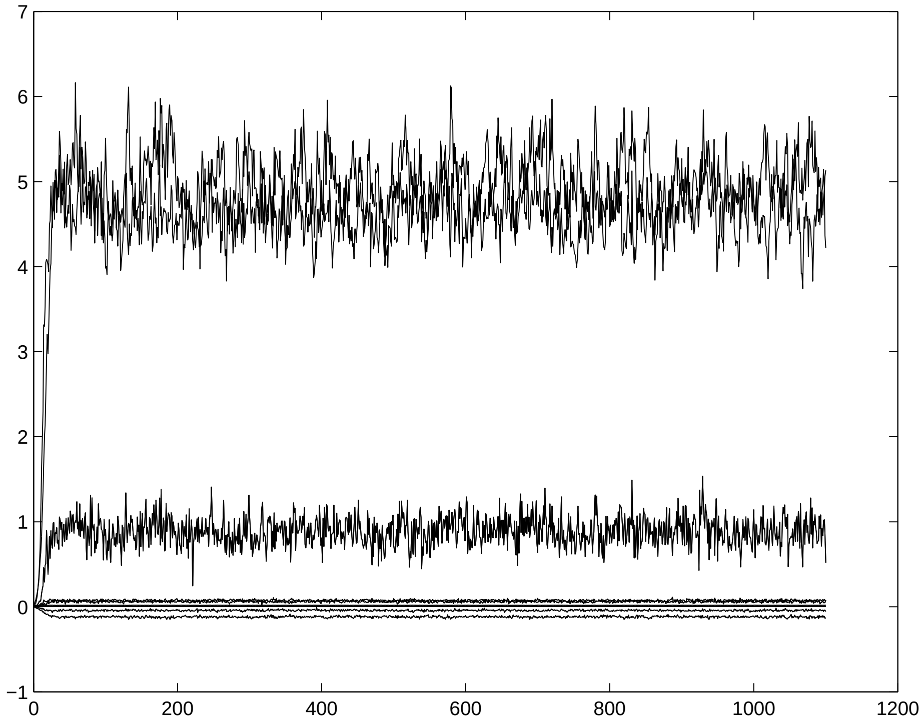
	number of cars	number of trucks	average car mile	average truck mile
number of cars	0.3007	-0.1342	1.7171	-2.5428
number of trucks	-0.1342	0.3245	-2.0854	3.2028
average car mile	1.7171	-2.0854	79.2414	-35.3766
average truck mile	-2.5428	3.2028	-35.3766	98.1505

4.1 Residential density

The density coefficient in the car choice equation is positive but insignificant; while the density coefficient in the truck choice equation is significantly negative at 1% level. This shows that the probability of choosing zero truck increases, while that of choosing two or more trucks decreases when there is an increase in residential density. To make quantitative statement of the impact in density increase, I calculated the change in probabilities when residential density increases by 1000 housing units per square mile.

From Table 4.1, we can see that the probability of not holding trucks increases by .0203, when density increases by 1000. In the extreme case, imagine a typical household moves from rural area of Sacramento, where the density is 150 housing units per square mile, to downtown LA, where the density is 6000 housing units per square mile. Its odd of not

Figure 1: MCMC draws of unique elements in inverse of the covariance matrix



holding trucks would increase by almost 12 percent, through which the odd of holding one truck would drop by around 8 percent, and that of holding two or more trucks by around 4 percent. The marginal effect of residential density on holdings of cars are insignificant yet positive. The positiveness can be explained as a substitution effect between cars and trucks when households move from low density areas to high density areas. When trucks are too costly to hold in high density areas, households might substitute trucks for more cars. The insignificance could imply that instead of substitution, some people might just reduce the total stock of vehicle within a household when moving to high density areas. The overall effect on the holding of cars would be ambivalent.

In terms of vehicle usage, an increase of residential density by 1000 will decrease the truck usage by 529.7 miles per year.

Table 4: Marginal effect of density

equation	probability changes in		
	Pr(n=0)	Pr(n=1)	Pr(n=2)
number of cars	-0.0034	-0.0034	0.0068
	(0.0036)	(0.0037)	(0.0066)
number of trucks	0.0203	-0.0141	-0.0062
	(0.007)	(0.0057)	(0.0033)

Notes: standard errors are reported in parentheses

4.2 Income

We see that the total annual income of the household has a significant association with choice and usage of trucks. However, since the income is represented by a series of dummies, it is difficult to get an sense of how big the effect of income is quantitatively. Here, I ran an empirical distribution test to get a smooth distribution function covering the histrogram of the original categorical income data. By drawing from a truncated normal distribution according to each category, pseudo income data are obtained. By simulation, a 10% increase in annual income will increase the probability of choosing two or more cars by 0.003, and that of choosing one or more trucks by 0.0154; a 50% increase in annual income will increase the probability of choosing two or more cars by 0.0255, and that of choosing one or more trucks by 0.0569. When income is doubled, these two numbers increases to 0.0414 and 0.0882.

Table 5: Income effects

	sample data	income increase 10%	income increase 50%	income increase 100%
Pr(car=0)	0.0329	0.0324	0.0297	0.0268
Pr(car=1)	0.6622	0.6597	0.6399	0.6269
Pr(car=2)	0.3049	0.308	0.3304	0.3463
Pr(truck=0)	0.5258	0.5104	0.4689	0.4376
Pr(truck=1)	0.4604	0.4747	0.5126	0.5413
Pr(truck=2)	0.0138	0.0148	0.0185	0.0211

Households with annual income between $20,000 and $30,000 tend to drive 4358.6 truck miles per truck annually more than those with annual income less than \$20,000. With increase in income, the average miles driven by trucks also tend to increase, albeit not smoothly. If gasoline price are same across California, which is not a wild assumption, higher income households can tolerate higher operational cost and hence drive more.

4.3 Non-mobile transportation

Total number of bicycles a household own is used here as an indicator for alternative transportation. The estimation result shows that number of bicycles is positively correlated with number of vehicle a household have, significantly in number of trucks. One additional bicycle is associated with around 4% increase in probability of holding trucks, at 1% significance level, and with 570 more miles driven by trucks, at 5% significance level. This indicate that bikes are not substitutional transportation mode of cars or trucks, but are instead complimentary. Understandably, many families use bikes for recreational uses and use trucks to carry bikes to recreational areas such as mountains, parks, and scenery bike lanes. Urban policy in increasing bike lanes in the hope of reducing mobile transportation usage might need to think twice.

4.4 Other variables

Working status of the respondent of a household also explains some different behaviors in vehicle choice and vehicle usage. Retired people tend to hold cars instead of trucks, and if they have trucks, tend to drive much (3652 miles per year) less than employed people. Homemakers exhibit similar behavioral patterns as retired people. Education is also an important factor affecting choice of vehicles. The more education the respondent receives, the more likely that household chooses cars over trucks. Holding other variables at their sample mean, a bachelor degree holder is 14% less likely to have trucks than a high school degree holder. The predictive probability of having zero trucks for a high school degree holder is .49, while that for a bachelor's degree is .63. High school dropouts has the highest probability of owning trucks among the three educational level recipients. They have almost 22% more probability to own trucks than people who have a bachelor's or higher degree.

Table 6: Educational effect

	Pr(c=0)	Pr(c=1)	Pr(c=2)	Pr(t=0)	Pr(t=1)	Pr(t=2)
$<$ highs chool	0.1219	0.6228	0.2553	0.4077	0.5283	0.0640
[high school bachelor)	0.0714	0.5686	0.3600	0.4866	0.4709	0.0425
$>$ bachelor	0.0562	0.5372	0.4066	0.6257	0.3554	0.0189

Possible explanations for the different behavior with different educational levels might due to environmental awareness. Higher education recipients may be more aware of the environmental damage caused by gasoline consumption. Or more probably, many high school dropouts are blue-collar workers that need to use trucks. In health economics, researchers have found that education and obesity are negatively correlated(Lahti-Koski1, Vartiainen, Mannisto, and Pietinen 2000), especially among women. This may provide another reason why lower education recipients tend to choose more spacious vehicles—trucks, than their higher education counterparts.

5 Conclusion

This paper provides a reduced-form discrete-continuous system to model households' vehicle holdings and usage decision. The system is composed of a bi-variate ordered probit model and a bi-variate tobit model. The ordered probit model is to capture households decision on number of vehicles in each category, here the categories are cars and trucks, for the purpose of environmental and energy saving policy implications. In using number of vehicles in each category instead of total number of vehicles, the analysis circumvents the usual difficulty that facing traditional modelling of vehicle holdings: with an increase in total number of vehicles, possible combination of vehicle holdings proliferates, the estimation then becomes too cumbersome to handle households with more than two vehicles. In this method that used in this paper, handling multiple vehicle households is simple and straightforward. The bi-variate tobit model is to capture households decision on miles driven conditional on each category. Traditional discrete-continuous models were built upon utility maximization in theory, but some crude approximation in estimation dull the elegance of theoretical basis. By uniting bi-variate ordered probit and tobit model and assume an unrestricted covariance matrix, we can cleanly estimate a reduced-form discrete/continuous system. By the aid of data augmentation and Bayesian Markov Chain Monte Carlo methods, the estimation is straightforward.

The estimation results show that the residential density has a negative statistically significant association with number of trucks held and miles driven. When density increases by 2000, the probability of holding no trucks increases by around 4 percent. This effect is relatively small but still comparable to effects caused by household demographic characteristics, for example, income effects or education effects, where a 50% increase in annual income will increase the probability of choosing two or more cars by 2.5 percent, and that of choosing one or more trucks by 5.7 percent; and people with a bachelor's or higher degree have 22% less probability to own trucks than highschool dropouts. Bicycles turn out be compliment to mobile transport instead of substitutive.

Appendix

Assume that

$$\beta \sim \mathscr{N}(\beta_0, V_0) \text{ and } \Sigma \sim \mathscr{IW}(\nu, Q^{-1})$$

The posterior distribution of the unknown parameters are:

$$\begin{aligned} f(\beta, \Sigma, \mathbf{y}^*|\mathbf{y}) &\propto f(\mathbf{y}, \mathbf{y}^*|\beta, \Sigma) f(\beta, \Sigma) = \prod_{i=1}^{T} f(y_i|y_i^*, \beta, \Sigma) f(y_i^*|\beta, \Sigma) f(\beta, \Sigma) \quad (18) \\ &\propto \prod_{i=1}^{T} |\Sigma^{-1}|^{1/2} \exp(-\frac{1}{2}(y_i^* - x_i\beta)'\Sigma^{-1}(y_i^* - x_i\beta)) I_{i=1,2}(\alpha_{y_i} < y_i^* \leq \alpha_{y_i+1}) \quad (19) \\ &\times \exp(-\frac{1}{2}(\beta - \beta_0)'V_0^{-1}(\beta - \beta_0) \times |\Sigma^{-1}|^{(\nu+k+1)/2} \exp(-\frac{1}{2} tr\Sigma^{-1}Q) \quad (20) \end{aligned}$$

Given the joint posterior distribution, the conditional posterior distributions are:

$$\begin{aligned} f(\beta|y_i, y_i^*, \Sigma) &\propto \exp(-\frac{1}{2}[\sum_{i=1}^{T}(y_i^* - x_i\beta)'\Sigma^{-1}(y_i^* - x_i\beta) + (\beta - \beta_0)'V_0^{-1}(\beta - \beta_0)]) \quad (21) \\ &\propto \exp(-\frac{1}{2}(\beta - \bar{\beta})'\bar{V}^{-1}(\beta - \bar{\beta})) \quad (22) \end{aligned}$$

where

$$\begin{aligned} \bar{V} &= (V_0^{-1} + \sum_{i=1}^{T} x_i'\Sigma^{-1}x_i)^{-1} \quad (23) \\ \bar{\beta} &= \bar{V}(V_0^{-1}\beta_0 + \sum_{i=1}^{T} x_i'\Sigma^{-1}y_i^*) \quad (24) \end{aligned}$$

Hence,

$$\beta|y_i, y_i^*, \Sigma \sim \mathscr{N}(\bar{\beta}, \bar{V}) \quad (25)$$

$$\begin{aligned} f(\Sigma|y_i, y_i^*, \beta) &\propto |\Sigma^{-1}|^{T/2} \exp(-\frac{1}{2} tr(\Sigma^{-1} \sum_{-1}^{T} (y_i^* - x_i\beta)'(y_i^* - x_i\beta))) && (26) \\ &\times |\Sigma^{-1}|^{(\nu+k+1)/2} \exp(-\frac{1}{2} tr\Sigma^{-1} Q) && (27) \\ &\propto |\Sigma^{-1}|^{(\nu+T+k+1)/2} \exp(-\frac{1}{2} tr\Sigma^{-1} (\sum_{i=1}^{T} (y_i^* - x_i\beta)'(y_i^* - x_i\beta) + Q)) && (28) \end{aligned}$$

Therefore,

$$\Sigma|y_i, y_i^*, \beta \sim \mathscr{IW}(\nu + T, (\sum_{i=1}^{T} (y_i^* - x_i\beta)'(y_i^* - x_i\beta) + Q)^{-1}) \quad (29)$$

$$f(\mathbf{y_i^*}|y_i, \beta, \Sigma) = f(y_{1i}^*|y_i, \beta, \Sigma) \cdot f(y_{2i}^*|y_{1i}^*, y_i, \beta, \Sigma) \cdot f(y_{3i}^*|y_{1i}^*, y_{2i}^*, y_i, \beta, \Sigma) \cdot f(y_{4i}^*|y_{1i}^*, y_{2i}^*, y_{3i}^*, y_i, \beta, \Sigma)$$

$$\begin{aligned} y_{1i}^*|y_i, \beta, \Sigma &= TN_{[\alpha_{y_i}, \alpha_{y_i+1}]}(\mu_1, \sigma_{11}) && (30) \\ y_{2i}^*|y_{1i}^*, y_i, \beta, \Sigma &= TN_{[\alpha_{y_i}, \alpha_{y_i+1}]}(\mu_{2|1}, \Sigma_{2|1}) && (31) \\ y_{3i}^*|y_{2i}^*, y_{1i}^*, y_i, \beta, \Sigma &= TN_{[-\infty, 0]}(\mu_{3|1,2}, \Sigma_{3|1,2}) && (32) \\ y_{4i}^*|y_{3i}^*, y_{2i}^*, y_{1i}^*, y_i, \beta, \Sigma &= TN_{[-\infty, 0]}(\mu_{4|1,2,3}, \Sigma_{4|1,2,3}) && (33) \end{aligned}$$

Table 7: Estimation results of the multivariate ordered probit and tobit model

Variable	Coefficient			
	car	truck	avg.cmile	avg.tmile
Residential density	.021	-.0598***	63.1	-529.7***
	(.0207)	(.0219)	(152.6)	(182.8)
Number of bikes	.0209	.1047***	-149	569.5**
	(.0293)	(.029)	(214.9)	(260)
Total number of people	.0435	-.029	875.6**	-376
	(.0551)	(.0552)	(410.8)	(494.8)
Number of adults	.6123***	.3287***	230.3	987.8*
	(.0739)	(.0743)	(528.7)	(635.3)
EMPLOYED	.0185	-.1323*	337.6	139.3
	(.09)	(.0977)	(690.9)	(840.8)
RETIRED	.1888**	-.4404***	291.6	-3513.1***
	(.1106)	(.1168)	(852.4)	(1073.2)
HOMEMAKE	-.01019	-.3294***	307.9	-2056.6**
	(.1182)	(.137)	(961.5)	(1155.9)
Income between 20k and 30k	-.0512	.1931*	-1184	4358.6***
	(.1193)	(.1358)	(941.5)	(1238.5)
Income between 30k and 50k	-.0872	.4878***	-515.7	4124***
	(.1025)	(.1171)	(812.2)	(1050.9)
Income between 50k and 75k	.0669	.6263***	-67.3	4865.9***
	(.1135)	(.1221)	(878.3)	(1156.1)
Income between 75k and 100k	.0686	.8193***	-713.7	6263.4***
	(.1275)	(.1335)	(986.2)	(1245.8)
Income greater than 100k	.2343**	.8756***	132	6025.6***
	(.1215)	(.1248)	(949.3)	(1209.1)
Owns home	.1333**	.2984***	-1725.3***	1153.2**
	(.0751)	(.0801)	(557.1)	(686.7)
MSA has rail	.1175*	-.0817	645.7	456.6
	(.0726)	(.0712)	(530.3)	(632.7)
Respondent completed highschool	.2996***	-.1998***	1258.4**	-2016.1***
	(.0796)	(.0849)	(593.3)	(736.1)
Respondent has bachelor's degree	.4217***	-.554***	1660.5***	-4061.5***
	(.848)	(.0853)	(634.1)	(792.6)
Youngest child under 6	.0832	.2131*	-377.5	2684.9**
	(.1455)	(.1437)	(1098.4)	(1286.1)
Youngest child between 6 and 15	.0818	.1099	-162.3	1647.9*
	(.1339)	(.1296)	(986.2)	(1162.2)
Youngest child over 21	.0703	-.1306	790.8	-319.4
	(.1494)	(.1494)	(1169.1)	(1290.5)
Urban	.1635*	-.4133***	1105.9	-3283.1***
	(.1158)	(.1144)	(875.6)	(1031.1)
Second city	.148*	-.2422**	544.1	-722.5
	(.1118)	(.1081)	(835.2)	(984.9)
Suburban	.2089**	-.3657***	1454.5**	-2091.4**
	(.1072)	(.1019)	(757.9)	(926.6)

Notes: standard errors are reported in parentheses below the estimates. *, **, and *** indicate statistical significance at 10, 5, and 1% levels, respectively.

6 References

Albert, J. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88, 669–679.

Brownstone, D. and Ken Small(1989), "Efficient Estimation of Nested Logit Models." *Journal of Business and Economic Statistics*, Vol. 7, No.1, 67-74

Berkovec, J., and J. Rust(1985), "A Nested Logit Model of Automobile Holdings for One Vehicle Households." *Transportation Research, Part B—Methodological*, 19, 275-285

Brueckner, J.(2001), "Urban Sprawl: Lessons from Urban Economics." in W. Gale and J. Pack, eds. Brookings-Wharton Papers on Urban Affairs.

Brueckner, J. and Ann G. Largey (2006), "Social Interaction and Urban Sprawl." Working paper

Cervero, Robert and Kara Kockelman(1997), "Travel Demand and the 3Ds: Density, Diversity and Design." *Transportation Research Part D: Transport and Environment*, 3, 199-219.

Dunphy, Robert T. and Kimberly Fishe(1996), "Transportation, Congestion, and Density: New insights." *Transportation Research Record*, 1552, 89-96.

Feng, Ye, Don Fullerton and Li Gan(2005), "Vehicle Choices, Miles Driven and Pollution Policies." NBER Working Paper Series, Working paper 11553.

Golob, F. Thomas, and David Brownstone (2005), " The Impact of Residential Density on Vehicle Usage and Energy Consumption. " Working paper, University of California, Irvine.

Small, Ken, and Ian W.H. Parry(2005),"Does Britain or The United States Have the Right Gasoline Tax" *American Economic Review*, 95, pp. 1276-1289

Lahti-Koski1, M., E. Vartiainen, S. Mannisto, and P. Pietinen (2000), "Age, education and occupation as determinants of trends in body mass index in Finland from 1982 to 1997." *International Journal of Obesity*, 24, 1669-1676

Mannering, Fred and Clifford Winston(1985), "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization." *The RAND Journal of Economics*, Vol.16, No.2, 215-236.

Nandram, Balgobin and Ming Hui Chen(1996), "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence." *Journal of Statistical Computation and Simulation*, Vol.54, 129-144

Train, Kenneth E. (2003), Discrete Choice Methods with Simulation. Cambridge, UK: Cambridge University Press.

West, Sarah(2004), "Distributional Effects of Alternative Vehicle Pollution Control Policies." *Journal of Public Economics*, 88, 735-757.