Quantifying the Contribution of Various Factors to Household Vehicle Miles of Travel

Abhilash Singh[1], Sebastian Astroza[1], Venu Garikapati[2], Ram Pendyala[3], Chandra Bhat[1,4] and Patricia Mokhtarian[5]

[1]The University of Texas at Austin, [2]National Renewable Energy Laboratory, [3]Arizona State University,
[4]The Hong Kong Polytechnic University, [5]Georgia Institute of Technology

Introduction

- Increases in Vehicle Miles Travelled (VMT) are **associated with higher levels of congestion and delay**, energy consumption, emissions, and roadway crashes.
- Earlier research has examined the effects of various factors on household VMT. However, research to date has not adequately documented the ***relative* contributions of various factors** influencing household VMT.
- The focus of this analysis is on disentangling and quantifying **the relative contributions of the following factors** on household VMT: **(1)** Household and person socio-economic and demographic characteristics, **(2)** Built environment attributes, **(3)** Residential self-selection effects, **(4)** Socio-spatial dependency effects, and Unobserved factors.

Modeling Methodology

Nominal Unordered Variable

- Using a typical utility maximizing framework, and the utility for alternative i and household q may be written as

$$U_{qi} = \boldsymbol{\beta}' \boldsymbol{x}_{qi} + \varepsilon_{qi}$$

- $\boldsymbol{x}_{qi}$ is a (K×1) column vector of **exogenous attributes,** $\boldsymbol{\beta}$ is a (K×1)-column vector of corresponding coefficients, ε_{qi} is a normal scalar error term.
- $\boldsymbol{U} \sim MVN_{QI}(\boldsymbol{V}, \mathbf{IDEN}_Q \otimes \boldsymbol{\Lambda})$, where $\mathbf{IDEN}_Q$ is an identity matrix of size Q.

Continuous Dependent Variable

- $y_q = \boldsymbol{\gamma}' \mathbf{z}_q + \eta_q$ is the usual regression equation, where the vector $\mathbf{z}_q$ of size C *x 1* includes a constant, exogenous variables, as well as dummy variables for each household location alternative. $\boldsymbol{\gamma}$ is the C *x 1* vector of coefficients.
- Let η_q be a **normally distributed idiosyncratic term** distributed independently and identically across households with mean zero and a variance of σ^2.
- Adding a **spatial dependence component** to the regression yields $y_q = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'} + \boldsymbol{\gamma}' \mathbf{z}_q + \eta_q$

  where $w_{qq'}$ are the the elements of an exogenously defined distance-based spatial /social weight matrix $\mathbf{W}$ corresponding to observations q and q' and δ is the spatial autoregressive parameter.
- The equation can thus be re-written as $\boldsymbol{y} = \mathbf{S}\boldsymbol{z}\boldsymbol{\gamma} + \mathbf{S}\boldsymbol{\eta}$
- Defining $\mathbf{S} = [\mathbf{IDEN}_Q - \delta \mathbf{W}]^{-1}$ we can write the above equation as $\boldsymbol{y} = \delta \mathbf{W}\boldsymbol{y} + \boldsymbol{z}\boldsymbol{\gamma} + \boldsymbol{\eta}$

Joint Model System

- The potential **endogeneity of residential choice** (that is, the self-selection of residence based on VMT desires) may be incorporated in the equations above by **allowing a covariance in the error terms** between the discrete and continuous dependent variables.
- The covariance matrix of the vector $\breve{\boldsymbol{y}}_q = (\breve{\boldsymbol{u}}_q, \eta_q)$ is defined as

  $\text{Cov}(\breve{\boldsymbol{y}}_q) = \begin{bmatrix} \breve{\boldsymbol{\Lambda}} & \boldsymbol{\Psi} \\ \boldsymbol{\Psi}' & \sigma^2 \end{bmatrix}$ where $\boldsymbol{\Psi}$ is an (I-1)x1 vector capturing covariance effects.
- The computation of the resulting likelihood function involves the **evaluation of a multi-dimensional integral** of the multivariate cumulative normal distribution, which is prohibitive even for medium-sized samples.
- Therefore, the **Maximum Approximate Composite Marginal Likelihood** (MACML) approach, in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is adopted for estimating the joint model.

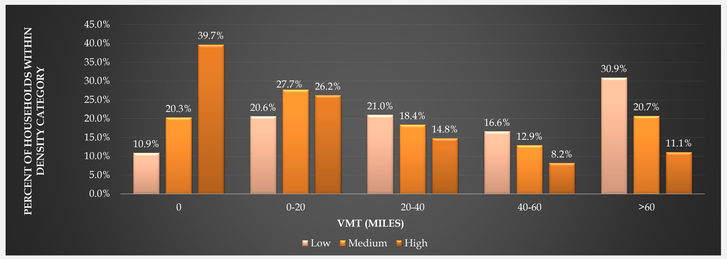
Data and Sample Description

- Data derived from the **2010-2011 Regional Household Travel Survey (RHTS) of the New York Metropolitan Transportation Council (NYMTC)**.
- Survey gathered information from **14,791 households.** A **random sample of 3000 households** was extracted for analysis purposes.
- Household VMT was computed by **aggregating distance traveled (in miles) across the personal vehicle trip records**, while explicitly ensuring that no trip was double-counted.
- The **density of the residential zone** was calculated by adding population and employment, and dividing the sum by the area of the zone.
- Each household was classified into a residential density category depending on whether it fell into the top third, middle third, or bottom third of zones ranked by land use density.

Table 1: Description of Survey Sample Used for Analysis

Dependent variable: Residential Location (Discrete) Variable				
Location density [(pop+emp) / area]	Number of observations (%)			
Low	1,000 (33.33)			
Medium	1,000 (33.33)			
High	1,000 (33.33)			
Dependent variable: Household VMT (Continuous) Variable				
Variable	Mean	Std Dev	Min	Max
Vehicle Miles Traveled (miles)	35.1	42.0	0	326.9
Natural log of Vehicle Miles Traveled	2.6	1.71	0	5.79
Residential Density Choice by Explanatory Variable				
	Low	Medium	High	Total
Family Structure Variables				
Single Person, N (%)	260 (28.2)	309 (33.5)	354 (38.4)	923 (100)
Single Parent, N (%)	26 (30.6)	31 (36.5)	28 (32.9)	85 (100)
Couple, N (%)	320 (38.9)	257 (31.2)	246 (29.9)	823 (100)
Nuclear Family, N (%)	201 (37.6)	182 (34.1)	151 (28.3)	534 (100)
Joint Family, N (%)	193 (30.4)	221 (34.8)	221 (34.8)	635 (100)
Total	1000	1000	1000	
Household Income Variables [US$/year]				
Below 30,000, N (%)	135 (23.3)	219 (37.8)	226 (39.0)	580 (100)
30,000 to 75,000, N (%)	283 (31.2)	311 (34.3)	313 (34.5)	907 (100)
>75,000 to 150,000, N (%)	381 (36.8)	324 (31.3)	330 (31.9)	1035 (100)
Above 150,000, N (%)	201 (42.1)	146 (30.5)	131 (27.4)	478 (100)
Total	1000	1000	1000	
Household race and ethnicity				
Caucasians, N (%)	788 (36.4)	719 (33.2)	659 (30.4)	2166 (100)
African American, N (%)	72 (20.7)	131 (37.8)	144 (41.5)	347 (100)
Hispanic, N (%)	36 (16.2)	84 (37.8)	102 (45.9)	222 (100)
Asian and other, N (%)	104 (39.2)	66 (24.9)	95 (35.8)	265 (100)
Total	1000	1000	1000	
Household Unit type				
Villa Detached Residence, N (%)	650 (41.9)	522 (33.7)	379 (24.4)	1551 (100)
Villa Attached Residence, N (%)	81 (36.3)	71 (31.8)	71 (31.8)	223 (100)
Condo Residence, N (%)	269 (21.9)	407 (33.2)	550 (44.9)	1226 (100)
Total	1000	1000	1000	

Figure 1. Distribution of Households in Each Density Category by VMT Class



Estimation Results

- Several spatial dependency forms were tested. Every specification that was attempted yielded an **insignificant spatial dependency effect** in the model.
- An independent model that ignores self-selection effects was also estimated. Results for the independent model are quite similar to those in the model with self-selection.
- In the model with self-selection, **significant error covariances exist between** households residing in **medium or high density neighborhoods and vehicle miles of travel**.

Table 2: Joint Residential Location (Density) and Aspatial Household VMT Model with Self-Selection

Variables	MNP Residential Choice			Continuous LR
	Low Density Coef (t-stat) (base)	Medium Density Coef (t-stat)	High Density Coef (t-stat)	ln (vehicle miles traveled) Coef (t-stat)
Constant	-	-0.1233 (-4.23)	-0.1929 (-5.37)	0.8429 (8.4)
Family structure Variables				
Single Person	-	-	0.1839 (3.62)	-
Couple	-	-	-	-
Nuclear Family	-	-	-	-
Joint Family	-	-	-	-
Household Income [US$/year]				
Below 30,000	-	0.2145 (3.15)	0.2069 (2.83)	-
30,000 to 75,000	-	-	-	-
75,000 to 150,000	-	-	-	-
Household race and ethnicity				
African-American	-	0.3342 (3.96)	0.4100 (4.84)	-
Hispanic	-	0.4533 (4.14)	0.6362 (5.85)	-
Other races	-	-	-	-
Fractions of hh in age-groups				
Age 16 to 35	-	-	0.1701 (2.01)	-
Age 35 to 55	-	-	-	0.2330 (3.13)
Age 55 to 65	-	-	-	0.2013 (2.73)
Age above 65	-	-	-	-
Residential Density				
Medium density	-	-	-	-0.4309 (-7.52)
High density	-	-	-	-0.7619 (-13.28)
Vehicles in household				
One vehicle	-	-	-	1.6606 (22.35)
Two or more vehicles	-	-	-	2.5955 (32.45)
Workers in household (count)	-	-	-	0.1505 (4.70)
Students in household (count)	-	-	-	0.1388 (2.55)
Fraction of unemployed in household	-	-	-0.3073 (-3.54)	-

Goodness of fit: The adjusted composite likelihood ratio test statistic for the joint model compared to that of the independent model is 6.18, which is larger than the critical χ^2 value with two degrees of freedom at 95% confidence level.

Conclusions and Future Research

- Based on the model estimation results, the **relative contributions of various factors** influencing household VMT are:

Socio-economic and demographic characteristics → 38.4%	**52.8%**	**Self-selection effects → 5.9%**
Built environment attributes → 8.5%		**Socio-spatial dependence → Insignificant**

- This leaves **47.2 percent of the variance in household VMT unexplained** by the factors considered in the model specification of this paper.
- Within the 52.8 percent of household VMT variance, the **socio-economic and demographic characteristics account for 72.5 percent** of the explained portion, residential self-selection accounts for 11.5 percent, and built environment attributes account for 16 percent.
- **Future research** should focus on adding more built environment information (such as proximity to transit infrastructure, and land use diversity) in the model specification to **further explore the impact of built environment on VMT**.