Transportation Research Part D 63 (2018) 23–36

Contents lists available at ScienceDirect

Transportation Research Part D

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

Quantifying the relative contribution of factors to household vehicle miles of travel

Abhilash C. Singh[a], Sebastian Astroza[a,b], Venu M. Garikapati[c], Ram M. Pendyala[d], Chandra R. Bhat[a,e,*], Patricia L. Mokhtarian[f]

[a] *The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, United States*
[b] *Departamento de Ingeniería Industrial, Universidad de Concepción, Edmundo Larenas 219, Concepción, Chile*
[c] *National Renewable Energy Laboratory, Systems Analysis & Integration Section, 15013 Denver West Parkway, Golden, CO 80401, United States*
[d] *Arizona State University, School of Sustainable Engineering and the Built Environment, 660 S. College Avenue, Tempe, AZ 85287-3005, United States*
[e] *The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[f] *Georgia Institute of Technology, School of Civil and Environmental Engineering, 790 Atlantic Drive, Atlanta, GA 30332-0355, United States*

ARTICLE INFO

Keywords:
Vehicle miles of travel (VMT)
Demographic effects
Built environment effects
Residential self-selection
Social-spatial dependence

ABSTRACT

Household vehicle miles of travel (VMT) has been exhibiting a steady growth in post-recession years in the United States and has reached record levels in 2017. With transportation accounting for 27 percent of greenhouse gas emissions, planning professionals are increasingly seeking ways to curb vehicular travel to advance sustainable, vibrant, and healthy communities. Although there is considerable understanding of the various factors that influence household vehicular travel, there is little knowledge of their *relative* contribution to explaining variance in household VMT. This paper presents a holistic analysis to identify the relative contribution of socio-economic and demographic characteristics, built environment attributes, residential self-selection effects, and social and spatial dependency effects in explaining household VMT production. The modeling framework employs a simultaneous equations model of residential location (density) choice and household VMT generation. The analysis is performed using household travel survey data from the New York metropolitan region. Model results showed insignificant spatial dependency effects, with socio-demographic variables explaining 33 percent, density (as a key measure of built environment attributes) explaining 12 percent, and self-selection effects explaining 11 percent of the total variance in the logarithm of household VMT. The remaining 44 percent remains unexplained and attributable to omitted variables and unobserved idiosyncratic factors, calling for further research in this domain to better understand the relative contribution of various drivers of household VMT.

1. Introduction

Vehicle miles of travel (VMT), a key measure of travel demand, is on the rise in the United States and countries around the world (Bastian et al., 2016; Polzin, 2016). Predictions of the peaking of travel, largely made during the period of the great recession, are

* Corresponding author at: The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, United States.
E-mail addresses: acsingh@utexas.edu (A.C. Singh), sastroza@utexas.edu (S. Astroza), venu.garikapati@nrel.gov (V.M. Garikapati), ram.pendyala@asu.edu (R.M. Pendyala), bhat@mail.utexas.edu (C.R. Bhat), patmokh@gatech.edu (P.L. Mokhtarian).

https://doi.org/10.1016/j.trd.2018.04.004

proving to have been premature (Polzin, 2016). While there are signs of some shifts in residential location and travel choices, most notably related to the lower levels of vehicle ownership and mobility depicted by millennials and a move towards urban living among different generations (Badger, 2014; Logan, 2014), the fact of the matter is that total VMT has grown steadily in the United States since 2012 and has reached record levels in 2017 even after accounting for population and employment growth (Economic Research, 2017). Increases in VMT are associated with higher levels of congestion and delay, energy consumption and greenhouse gas emissions, and roadway crashes (Sacramento Area Council of Governments, 2016) – adversely affecting human health, quality of life, and community resiliency and sustainability (Levy et al., 2010). The growing presence of transportation network companies that provide mobility-as-a-service and the potential advent of autonomous vehicles may further contribute to an increase in VMT as travel becomes increasingly convenient and less burdensome, thus resulting in a reduced value of travel time.

For the reasons noted above, planning professionals in cities around the world are continuously seeking ways to reduce vehicle miles of travel without inhibiting household and business activity engagement. Formulating policies, strategies, and transportation infrastructure improvements that would reduce VMT is difficult, however, in the absence of an accurate understanding of the contribution of various factors to total VMT. This paper aims to provide a comprehensive understanding and quantification of the relative effects of various factors on household vehicle miles of travel. The analysis focuses on household VMT because it constitutes more than 75% of total VMT in the United States (AASHTO, 2013), and hence strategies aimed at curbing household VMT would likely yield the most benefits to communities.

There is undoubtedly an abundance of research that has examined the effects of various factors on household VMT in various geographic contexts (e.g., Millard-Ball and Schipper, 2011; Bastian et al., 2016). However, research to date has not adequately documented the *relative* contribution of various factors to explaining household VMT, thus calling for a more holistic and comprehensive analysis that is capable of doing so. While some studies explain the effects of socio-economic and demographic characteristics on VMT, others focus on examining the effects of built environment attributes on VMT. These studies are undoubtedly valuable, but it is also important to quantify the relative contribution of different factors to household VMT. By doing so, it is envisioned that planners and policy makers will be able to develop targeted policies that more effectively reduce vehicular travel. If, for example, built environment attributes are found to explain the variation in household VMT more than other factors (such as socio-economic and demographic factors), then decision-makers may realize the most benefits (in terms of VMT reductions) by implementing policies that foster more walkable, dense, and diverse built environments. On the other hand, if social interaction and spatial dependency effects are found to contribute more heavily to explaining variance in household VMT (relative to other factors), then policy makers may be well served by focusing resources on social media and public information campaigns that would facilitate spread of awareness (say, about use of alternative modes of transportation) through network diffusion mechanisms. While literature provides some information about the effects of these factors when viewed independently or in pairs, there is a lack of research dedicated to explaining the relative contribution of various factors in a comprehensive framework. This research effort is aimed at addressing this critical gap in the existing literature. Not only does this paper aim to offer insights on the relative contribution of various factors to household VMT, but the paper also aims to offer a rigorous methodological framework that is generalizable and can be applied in any geographical context. Thus this study is motivated by both methodological and empirical objectives with a view to help advance the development of sustainable communities.

This paper considers four different factors that may explain the variance in household VMT. These include household and person socio-economic and demographic characteristics, residential built environment attributes, residential self-selection (i.e., lifestyle preference) effects, and human social and spatial dependency effects. As noted earlier, while there are a number of research efforts that have examined the effects of subsets of these factors on household or personal VMT, there is no study that examines the relative contribution of each of these effects on household VMT in a singular holistic framework. The four factors considered in this paper are those that have been shown to influence household VMT in significant ways. Household socio-economic and demographic characteristics, such as household size, number of children, number of workers, and household income affect household VMT. Built environment attributes including land use density, population and employment density, parking availability and pricing, distance from residence to work centers, and multimodal accessibility (to destinations) affect household VMT. Residential self-selection effects capture the notion that individuals may choose to locate (live and work) in built environments that are consistent with their attitudes (e.g., environmental sensitivity) and lifestyle preferences (e.g., car-free lifestyle). The fourth and final factor considered in this paper is the socio-spatial dependence effect. Household VMT may be shaped by social interaction and spatial dependency effects, capturing influences engendered by people's interactions and geographic proximity. It should be noted that, even after accounting for these four factors, a residual unexplained effect will inevitably exist.

The analysis in this paper is performed on the 2010–2011 Regional Household Travel Survey (RHTS) of the New York Metropolitan Transportation Council (NYMTC). From the fall of 2010 through the fall of 2011, travel data was collected from 19,000 households across 28 counties in New York, New Jersey, and Connecticut (New York Metropolitan Transportation Council, 2011). After merging built environment data with the travel survey records, a joint model of residential location (density) choice and household VMT – accounting for residential self-selection and socio-spatial dependency effects – is estimated to unravel the relative contribution of various factors in explaining variance in household VMT.

The remainder of the paper is organized as follows. The next section presents a brief discussion of factors that influence household VMT. The third section presents a data description, the fourth section offers a description of the methodology, and the fifth section presents model estimation results. The sixth and final section offers a discussion and interpretation of the results together with concluding thoughts.

2. Explaining household vehicle miles of travel

Exploring factors that influence household and person VMT has been a topic of considerable interest for several decades, largely due to the contribution of VMT to traffic congestion, emissions, and energy consumption. Cervero and Kockelman (1997) used data from the 1990 San Francisco Bay Area travel survey to examine the role of built environment characteristics in shaping VMT and mode choice. They found that density, land use diversity, and pedestrian-oriented designs reduce trip rates, and encourage non-motorized mode use. More recently, Zhang et al. (2012) re-examined the relationship between land use and VMT using data from five metropolitan areas in the US. In addition to corroborating earlier findings, they identify urban area size, status of the existing built environment, transit service coverage and service quality, and land use decision-making processes as major factors that influence household VMT. Based on data from 370 urbanized areas in the United States, Cervero and Murakami (2010) found that population size is significantly positively correlated with VMT per capita. Krizek (2003) studied changes in travel behavior that result from changes in neighborhood accessibility and concluded that relocating to areas with high accessibility decreases household VMT. Based on a meta-analysis of the literature on built environment and travel behavior, Ewing and Cervero (2010) conclude that VMT is most strongly influenced by accessibility to destinations. You et al. (2014) estimate a model to predict the total motorized mileage of a household based on various socio-demographic, built environment, and network accessibility measures. Not only do they find that socio-economic characteristics influence household VMT, but they also find that zonal accessibility to destinations is an important predictor of VMT. A number of studies have shown that there is a significant association between built environment attributes and non-motorized travel (walking and bicycling) (e.g., Frank and Engelke, 2001; Lee and Moudon, 2006; Copperman and Bhat, 2007; Cao, 2010).

In addition to exploring the role of observed covariates, a number of studies have attempted to account for self-selection effects when examining the influence of various attributes on household VMT. Brownstone and Golob (2009) used the California subsample of the 2001 National Household Travel Survey to estimate a joint model of residential density, vehicle use, and fuel consumption that takes residential self-selection effects into account. They infer that an increase in density of 1000 dwelling units per square mile in a zone equates to a decrease of 1200 VMT per year for a representative household. Using a quasi-longitudinal design that takes self-selection effects into account, Handy et al. (2005, 2006) studied the relationship between neighborhood characteristics and travel behavior. They report that built environment attributes significantly impact travel behavior, even after accounting for the effects of neighborhood self-selection.

Several studies have attempted to unravel the extent to which different factors contribute to variance in vehicular travel, but do so in the context of examining the influence of one or two factors at a time. For example, Zhou and Kockelman (2008) used a sample selection model, and find that self-selection accounts for anywhere between 10 and 42% of the total influence of the built environment on VMT. Bhat and colleagues (see, for example, Bhat and Guo (2007), Pinjari et al. (2009), and Bhat et al. (2016)) present methodologies to control for self-selection effects, and apply their frameworks to study the effects of built environment attributes on residential location choices and time-use/mobility-related decisions. They find that self-selection contributes anywhere from 4% to 58%, depending upon the precise time-use/mobility-related choice dimension being examined. Both Cao and Fan (2012) and Bhat et al. (2014) find that self-selection accounts for 28% of the overall built environment effect, while the remaining 72% constitutes the true built environment effect. In a recent study using data from the Greater Salt Lake region, Ewing et al. (2016) report that the (direct and total) effects of the built environment on VMT is about twice as much as the residential self-selection effect.

Other studies have explored the role of spatial dependency effects in shaping variables that influence household VMT (though, to our knowledge, the current paper is the first to directly consider spatial effects in the context of household VMT). As identified by Bhat et al. (2017), there has been recognition in the travel behavior literature that household and individual travel decisions are influenced by spatial interaction effects and social group effects (through a peer effect or a peer pressure effect) inside urban communities (Salvy et al., 2009; Ferdous et al., 2011). For example, Dill and Voros (2007) found that if an individual's co-workers bicycle to work, the individual is more likely to bicycle to work too. The notion of norms in one's social or neighborhood group impacting bicycling behavior is also consistent with the theory of planned behavior and the theory of interpersonal behavior (see Heinen et al., 2010). As another example of earlier travel behavior studies that consider spatial/social interactions, Adjemian et al. (2010) investigate the spatial inter-dependence in vehicle type choice using data from the 2000 San Francisco Bay Area Travel Survey and conclude that spatial dependence effects are significant in explaining the ownership of nearly every vehicle body type in the study region. Similarly, Paleti et al. (2013a) use a multinomial probit formulation that incorporates spatial interaction effects in the analysis of household vehicle fleet composition. They use mean distance between households to capture the spatial dependence effect, and find that spatial dependency plays a significant role in explaining vehicle acquisition choices. McDonald (2007) analyzes the association between neighborhood social environment and children's decision to walk to school, and finds evidence that parental perception of neighborhood cohesion greatly influences the decision of children walking to school. Bhat et al. (2017) find significant residential location-based spatial dependence in their analysis of individual-level bicycling frequency, using data from the 2014 Puget Sound Household Travel Survey, while Bhat et al. (2010) and Sener and Bhat (2012) similarly observe spatial dependency effects in the context of individual daily activity participation using the 2000 San Francisco Bay Area Travel Survey.

This illustrative review of the literature reveals the emergence of at least four factors that are potentially key determinants of household VMT. While past research in this domain has examined the effects of different attributes on household VMT in isolation from one another, this paper aims to quantify the *relative* contribution of each of these effects on household VMT, and thus contributes significantly to better understanding the role of each factor in shaping VMT. Even after accounting for these four factors, there will inevitably be a remaining unexplained portion of household VMT variance. The size of this portion is estimated as well.

3. Data and sample description

The data used in this study is derived from the 2010–2011 Regional Household Travel Survey (RHTS) conducted by the New York Metropolitan Transportation Council (NYMTC) and the North Jersey Transportation Planning Authority (NJTPA). The RHTS collected travel information for each household resident in the sample for one weekday. After extensive data cleaning, the household level data set included information for 14,791 households that provided complete information on a host of socio-economic, demographic, location, and travel variables of importance to this study. The sample contains households residing in the New York metropolitan region, including parts of the States of New Jersey and Connecticut.

The dependent variable of interest in this paper is *weekday household vehicle miles of travel (VMT)*, largely because this measure can be obtained from most household travel survey data sets. Trip records provided by individual household members were used to derive VMT estimates at the household level. Household VMT is defined in this paper as being exclusively based on trips that are made by personal vehicle only. The 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 in the VMT calculation*. Thus, for example, if two household members travel together, only the mileage associated with the trip reported by the driver is counted towards calculating VMT. This was done to ensure that a clear distinction is drawn between vehicle miles of travel (VMT) and person miles of travel (PMT), and focus the analysis in this paper exclusively on *household-level* VMT, which is naturally influenced by the extent to which household members travel jointly (rideshare or carpool). After calculating household VMT and appending the value to household records, data describing the traffic analysis zone (TAZ) of residence was also joined to the data set. Households were geo-located at the TAZ level, and data describing population and employment characteristics of the residence TAZ could be easily appended to the household travel survey data set.

For the current study, a random sample of 3,000 households was extracted for analysis purposes. Comparisons were performed to ensure that the random sample is representative of the original sample of 14,791 households. A smaller random sample was chosen for analysis purposes to avoid the risk of inflated test statistics that often accompanies model estimation with large sample sizes. A sample size of 3,000 households was considered sizeable enough to obtain reliable parameter estimates while avoiding artificially inflated test statistics that could lead to erroneous inferences. Table 1 provides an overview of the descriptive characteristics of the sample. The density of the residential zone was calculated by adding population and employment, and dividing the sum by the area of the zone. Then, each household was classified into a residential-zone density category depending on whether it fell into the top third, middle third, or bottom third of zones ranked by land use density.

The use of residential-zone density as the sole descriptor of the built environment is not without reason. In general, density is a measure that is easy to quantify, understand, and interpret. There are many other measures of built environment, but they are not necessarily as well-defined and quantified. Measures such as walkability index, pedestrian-friendliness, transit connectivity and service, land use diversity, and access to destinations are appealing, but not as easily defined. Density has been used extensively to characterize the built environment (e.g., Kim and Brownstone, 2013; Paleti et al., 2013b; Cao and Fan, 2012; and Bhat et al., 2016). In addition, using density alone as the built environment measure allows a clean identification of built environment and residential self-selection effects in explaining household VMT without problems of multi-collinearity of density with other built environment characteristics. The reader is also referred to an online supplement at http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/HouseholdVMT/OnlineSupplement.pdf discussing this issue in some more detail. In the rest of this paper, we will use the terms density and the built environment interchangeably, for ease in presentation.

Average household VMT is found to be 35 miles; an examination of the distribution of VMT showed that 23.6 percent of the households had zero VMT (which means the household members made absolutely no personal vehicle trips; for our empirical analysis, the VMT for these households is assigned a nominal value of 1 mile, so that the logarithm of VMT, used as the dependent variable in the modeling, is zero). 20.9 percent of the households had VMT equal to or greater than 60 miles. In this research study, residential location (density) choice (three-category discrete dependent variable) and household VMT are modeled jointly to unravel the contribution of various effects of interest. The remainder of Table 1 provides descriptive statistics for a few socio-economic variables. Among single persons, the largest percentage reside in high density areas; the opposite is true for couples and nuclear families. Among households with low income, the largest share resides in higher density zones. As income increases, the proportions shift, with the larger shares seen in the lower density zones. This is consistent with expectations that higher income individuals seek to reside in suburban lower density areas characterized by good schools, safe neighborhoods, open spaces, and larger homes. Among Caucasians, the largest percentages reside in low density neighborhoods. Minority households show an opposite pattern, with larger percentages residing in high density neighborhoods. Other descriptive results are also as expected.

Fig. 1 depicts the distribution of households in each land use density category by VMT. The figure shows an overall pattern that is consistent with expectations. For example, only 10.9 percent of households in the low density category report zero VMT, while 39.7 percent of households in the high density category report zero VMT. Conversely, 30.9 percent of households in the low density category report more than 60 vehicle miles of travel; but only 11.1 percent of high density households do so. For households in the high density category, the percent of households reporting higher levels of VMT drops noticeably (except for a slight anomaly in transitioning between the 40–60 mile range and the > 60 mile range). The overall patterns are quite discernible and consistent with expectations that households in higher density locations generate fewer VMT, possibly due to greater access to destinations and alternative modes of transportation. However, as noted in the literature, other effects are likely to be at play as well; households residing in different neighborhood densities differ with respect to socio-economic and demographic characteristics and lifestyle preferences (leading to residential self-selection effects). In addition, there may be spatial dependency effects (i.e., households' behavior is shaped by their interactions with and observation of other households in geographic proximity) that shape household

Table 1
Descriptive characteristics of the analysis sample (N = 3,000 households).

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
Independent variables distribution				
Variable	Mean	Std Dev	Min	Max
Number of workers in household	1.25	0.86	0.00	6.00
Presence of students in household (dummy)	0.34	0.47	0.00	1.00
Fraction of unemployed in household	0.13	0.27	0.00	1.00
Residential-zone density choice by explanatory variable				
	Low	Medium	High	Total
Family structure variables	1000 (33.3)	1000 (33.3)	1000 (33.3)	3000 (100)
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)
Household income variables [US$/year]	1000 (33.3)	1000 (33.3)	1000 (33.3)	3000 (100)
Below 30,000, N (%)	135 (23.3)	219 (37.8)	226 (39)	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)
Household race and ethnicity	1000 (33.3)	1000 (33.3)	1000 (33.3)	3000 (100)
Caucasian, 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 (10)
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)
Household unit type	1000 (33.3)	1000 (33.3)	1000 (33.3)	3000 (100)
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)
Households with members in age-groups				
Age below 16, N (%)	247 (34.1)	247 (34.1)	231 (31.8)	725 (100)
Age 16 to 35, N (%)	286 (32.2)	300 (33.8)	302 (34.0)	888 (100)
Age 35 to 55, N (%)	501 (32.5)	527 (34.2)	515 (33.4)	1543 (100)
Age 55 to 65, N (%)	410 (34.3)	409 (34.2)	378 (31.6)	1197 (100)
Age above 65, N (%)	197 (33.9)	197 (33.9)	187 (32.2)	581 (100)
Total[a]	1641	1680	1613	4934
Number of vehicles in the household	1000 (33.3)	1000 (33.3)	1000 (33.3)	3000 (100)
Zero vehicles, N (%)	71 (12.3)	169 (29.3)	336 (58.3)	576 (100)
One vehicle, N (%)	291 (31.1)	320 (34.2)	324 (34.6)	935 (100)
Two or more vehicles, N (%)	638 (42.8)	511 (34.3)	340 (22.8)	1489 (100)

[a] A household can belong to more than one category; hence the columns do not necessarily add to 1,000 or 3,000.

VMT. The objective of this paper is to quantify the relative contributions of each of these factors to explaining household VMT.

It should be noted that the choice of New York as the region for analysis is based on a few key considerations. To unravel the contribution of different factors to explaining household VMT, it is desirable to analyze a geographic context where there is considerable heterogeneity in built environment attributes, transit service levels, and socio-economic characteristics. The New York metropolitan region offers rich variance in the dependent variable (household VMT) and explanatory factors of interest. In addition, with increasing levels of urbanization and challenges faced by large cities around the world, it was considered useful to analyze a large metropolitan context such as New York.

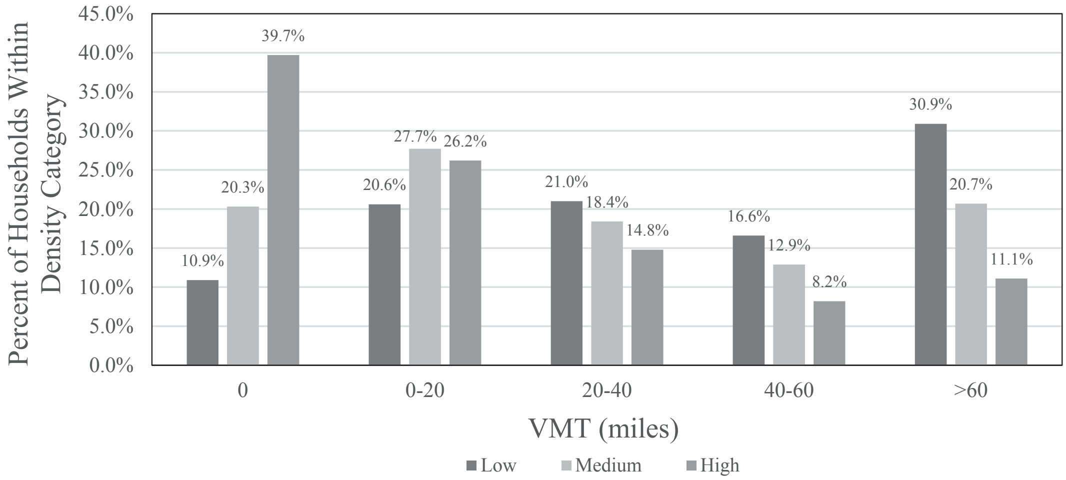

Fig. 1. Distribution of households in each density category by VMT class.

4. Modeling methodology

In this section, a brief overview of the modeling methodology is offered. The formulation for each variable is presented first, followed by a presentation of the structure and estimation procedure for the multi-dimensional model system of residential location (density) choice and household VMT production.

4.1. Nominal unordered variable (residential choice)

Let I ($I \geqslant 2$) be the number of alternatives corresponding to the nominal variable (residential location in the empirical analysis) and let i be the corresponding index (i = 1, 2, 3, …, I). Let Q be the number of households in the sample, and let q be the corresponding index (q = 1, 2, …, Q). Note that I may vary across households in a general discrete choice case, but the same number of alternatives is assumed across all households in this study. Using a typical utility maximizing framework, the utility for alternative i and household q may be written as:

$$U_{qi} = \boldsymbol{\beta}'\boldsymbol{x}_{qi} + \varepsilon_{qi}, \tag{1}$$

where $\boldsymbol{x}_{qi}$ is a $(K \times 1)$-column vector of exogenous attributes, $\boldsymbol{\beta}$ is a $(K \times 1)$-column vector of corresponding coefficients, and ε_{qi} is a normal scalar error term. Let the variance-covariance matrix of the vertically stacked vector of errors $\boldsymbol{\varepsilon}_q = [(\varepsilon_{q1},\varepsilon_{q2},...,\varepsilon_{qI})']$ be $\boldsymbol{\Lambda}$. The size of $\boldsymbol{\varepsilon}_q$ is $(I \times 1)$, and the size of $\boldsymbol{\Lambda}$ is $(I \times I)$. The error vector $\boldsymbol{\varepsilon}_q$ is identically and independently distributed across households. The model above may be written in a more compact form by defining the following vectors and matrices: $\boldsymbol{U}_q = (U_{q1},U_{q2},...,U_{qI})'$ $(I \times 1$ vector), $\boldsymbol{x}_q = (\boldsymbol{x}_{q1},\boldsymbol{x}_{q2},\boldsymbol{x}_{q3},...,\boldsymbol{x}_{qI})'$ $(I \times K$ matrix), and $\boldsymbol{V}_q = \boldsymbol{x}_q\boldsymbol{\beta}$ $(I \times 1$ vector). Then, $\boldsymbol{U}_q \sim MVN_I(\boldsymbol{V}_q,\boldsymbol{\Lambda})$, where $MVN_I(\boldsymbol{V}_q,\boldsymbol{\Lambda})$ is the multivariate normal distribution of I dimensions with mean vector $\boldsymbol{V}_q$ and covariance $\boldsymbol{\Lambda}$. Further, for future use, define $\boldsymbol{U} = (U_1',U_2',...,U_Q')'$ $(QI \times 1$ vector), $\boldsymbol{x} = (\boldsymbol{x}_1',\boldsymbol{x}_2',\boldsymbol{x}_3',...,\boldsymbol{x}_Q')'$ $(QI \times K$ matrix), $\boldsymbol{\varepsilon} = [(\varepsilon_1',\varepsilon_2',...,\varepsilon_Q')']$, $\boldsymbol{V} = \boldsymbol{x}\boldsymbol{\beta}$ $(QI \times 1$ vector), so that $\boldsymbol{U} \sim MVN_{QI}(\boldsymbol{V},\mathbf{IDEN}_Q \otimes \boldsymbol{\Lambda})$, where $\mathbf{IDEN}_Q$ is an identity matrix of size Q. Consider now that household q chooses alternative m. Under the utility maximization paradigm, $U_{qi}-U_{qm}$ must be less than zero for all $i \neq m$, since the household chose alternative m. Let $u_{qim} = U_{qi}-U_{qm} (i \neq m)$, and stack the latent utility differentials into an $[(I-1) \times 1]$ vector $\boldsymbol{u}_q = [(u_{q1m},u_{q2m},...,u_{qIm})';i \neq m]$. Also, let $\boldsymbol{u} = (\boldsymbol{u}_1',\boldsymbol{u}_2',...,\boldsymbol{u}_Q')'[Q(I-1) \times 1]$ vector. For future use, also define the utility differences with respect to the first alternative as $u_{qi1} = U_{qi}-U_{q1} (i \neq 1)$, $\breve{\boldsymbol{u}}_q = [(u_{q21},...,u_{qI1})']$, and $\breve{\boldsymbol{u}} = (\breve{\boldsymbol{u}}_1',\breve{\boldsymbol{u}}_2',...,\breve{\boldsymbol{u}}_Q')'[Q(I-1) \times 1]$.

In the context of the formulation above, several important identification issues (see Bhat (2015) for details) need to be addressed (in addition to the usual identification consideration that one of the alternatives has to be used as the base when introducing alternative-specific constants and variables that do not vary across the I alternatives). First, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}$ of $\breve{\boldsymbol{u}}_q$ are estimable (for each and all q). Thus, $\boldsymbol{\Lambda}$ is constructed from $\breve{\boldsymbol{\Lambda}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. Second, an additional scale normalization needs to be imposed on $\breve{\boldsymbol{\Lambda}}$. For this, we normalize the first element of $\breve{\boldsymbol{\Lambda}}$ to the value of one. Third, in MNP models, identification is tenuous when only household-specific covariates are used (see Keane (1992) and Munkin and Trivedi (2008)). In particular, exclusion restrictions are needed in the form of at least one household characteristic being excluded from each alternative's utility in addition to being excluded from a base alternative (but appearing in some other utilities).

4.2. Continuous dependent variable

In the empirical context of the current paper, the continuous variable is the natural logarithm of household vehicle miles of travel

(VMT). Let $y_q = \boldsymbol{\gamma}'\boldsymbol{z}_q + \eta_q$ in the usual linear regression fashion, where the vector $\boldsymbol{z}_q$ of size $(C \times 1)$ includes a constant, exogenous variables, and dummy variables for each household location alternative (except a base alternative).[1] $\boldsymbol{\gamma}$ is a corresponding $(C \times 1)$ 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. To the above equation, a spatial dependence component is now added using a typical spatial lag dependence specification as follows[2]:

$$y_q = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'} + \boldsymbol{\gamma}'\boldsymbol{z}_q + \eta_q \tag{2}$$

The $w_{qq'}$ terms are the elements of an exogenously defined distance-based spatial/social weight matrix $\mathbf{W}$ corresponding to observations q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$), and δ ($|\delta| < 1$) is the spatial autoregressive parameter. The weights $w_{qq'}$ can take the form of a discrete function such as a contiguity specification ($w_{qq'} = 1$ if the households q and q' are adjacent and 0 otherwise) or a specification based on a spatial/social distance threshold ($w_{qq'} = c_{qq'}/\sum_{q'} c_{qq'}$, where $c_{qq'}$ is a dummy variable taking the value 1 if the household q' is within the distance threshold and 0 otherwise). It can also take a continuous form such as those based on the inverse of distance $d_{qq'}$ and its power functions ($w_{qq'} = (1/\mathrm{d}_{qq'}^{n})\left[\sum_{q'} 1/\mathrm{d}_{qq'}^{n}\right]^{-1}$)($n > 0$), the inverse of exponential distance, and the shared edge length $\widetilde{d}_{qq'}$ between households (or observation units) $w_{qq'} = \widetilde{c}_{qq'}\widetilde{d}_{qq'}/(\sum_{q'} \widetilde{c}_{qq'}\widetilde{d}_{qq'})$ (where $\widetilde{c}_{qq'}$ is a dummy variable taking the value 1 if q and q' are adjoining based on some pre-specified spatial criteria, and 0 otherwise).[3]

Eq. (2) can be written equivalently in vector notation as:

$$\boldsymbol{y} = \delta\ \mathbf{W}\boldsymbol{y} + \boldsymbol{z}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{3}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_Q)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_Q)'$ are $(Q \times 1)$ vectors, $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_Q)'$ is a $(Q \times C)$ matrix of exogenous variables for all Q units, and, for future use, $\mathrm{Cov}(\boldsymbol{\eta}) = \mathbf{IDEN}_Q \otimes \sigma^2$. Defining $\mathbf{S} = [\mathbf{IDEN}_Q - \delta\ \mathbf{W}]^{-1}$ [$(Q \times Q)$ matrix], Eq. (3) may be re-written as:[4]

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}\boldsymbol{\gamma} + \boldsymbol{S}\boldsymbol{\eta} \tag{4}$$

4.3. The joint model system

The potential endogeneity of residential choice (that is, the self-selection of residence based on VMT desires) may be incorporated

[1] Even though we use the household location alternative as a determinant variable in the VMT equation, it is important to note that the system being estimated here is a joint model of residential location and VMT. That is, both residential location and VMT are endogenous variables (are co-determined) because, as we discuss later, we allow a correlation between the error vector $\breve{\boldsymbol{u}}_q$ in the residential location equation and the error term η_q in the VMT equation. Thus, any endogenous effect of residential location is a "true" causal effect after removing any spurious associations caused by neighborhood self-selection effects. For example, a household that is "green" may self-select into dense neighborhoods because of ease of access by walk to activity opportunities. If this "green" lifestyle is unobserved, it would result in a negative correlation between $\breve{\boldsymbol{u}}_q$ for the high density location alternative and the error term in the VMT equation. By recognizing and incorporating this correlation in a joint model of residential location choice and VMT (which reflects residential self-selection), we are then able to assess any "true" causal effects of high density living on household VMT (through the impact of the high density dummy variable in the VMT equation).

[2] The decision-making unit for analysis purposes is the household, but the density associated with a household corresponds to the zone of residence of the household.

[3] In the subsequent empirical analysis, the authors tested two different versions of the contiguity specification: a weight matrix based on defining "neighbors" of a household as other households in the same traffic analysis zone (TAZ), and an alternative weight matrix based on defining "neighbors" as other households in the same TAZ or an adjacent TAZ. All other weight matrix specifications listed in the main text were also tested.

[4] That spatial dependence is an important factor that influences travel behavior has already been established and recognized, as discussed earlier in the penultimate paragraph of Section 2. Further, as discussed by Anselin (2003) and Elhorst and Vega (2013), this spatial dependence may or may not include spatial spillovers (in which a change in an exogenous variable at one location affects the dependent variable at another location), and may or may not include global spillover effects (in which a change in an exogenous variable affects the dependent variable at all other locations, even though this effect fades with spatial separation). A complete discussion of these issues is beyond the scope of this paper. But suffice it to say that the spatial dependency assumed in this paper (which corresponds to the standard spatial lag model) corresponds to a global spatial spilllover effect as in Eq. (4). At the same time, the spatial lag model also allows spatial autocorrelation dependence in the unobserved component $\boldsymbol{\eta}$. Our reason for using the spatial lag model is based on two theoretical considerations rather than a data-driven statistical fit exercise. The first consideration, which we actually have not seen explicitly stated in the literature, is that the spatial lag model of Eq. (3) corresponds to the reduced form of Eq. (4), which itself may be written as $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{E}(\boldsymbol{y}) + \boldsymbol{S}\boldsymbol{\eta}$, or as $\boldsymbol{E}(\boldsymbol{y}) = \boldsymbol{S}\boldsymbol{E}(\boldsymbol{y})$. What this implies is that the global spillover applies to the expected values of the dependent variable. We argue that this is the correct way to model spillovers, given that the estimation is being done on but a sample of the population. That is, while social/spatial spillover effects happen through interactions and knowledge spillovers that influence a person's dependent variable behavior (VMT in the current context) based on other individuals' decisions on the dependent variable at other locations in the population (that is, through a relationship among the elements of the $\boldsymbol{y}$ vector in the population, as documented in Section 2), it is impossible to represent every individual in spatial or social space. Thus, a sampled individual at a particular location can only be considered as representative of many other non-sampled individuals at that location. Thus, it is intrinsically more appropriate to consider that, because we are using only a sample, the spillover effects operate on the expected values. The second consideration for using a spatial lag model is that we require symmetry in the spillover effects and the spatial autocorrelation generated through the unobserved individual-specific error terms. This is because, if for some reason, we drop an independent variable from the model so that the effect of the variable is moved from the substantive portion of the model to the unobserved portion (or if we add an independent variable so that part of the unobserved portion is now included with the observed portion), the implied spatial dependence should not change. The only spatial structure that guarantees this is the spatial lag model, not the other spatial dependence structures identified in the literature. Thus, on pure theoretical and logical grounds, spatial dependency in VMT models (and in travel behavior models in general) should be developed using the spatial lag formulation we use here. In essence, while some papers (see Pinkse and Slade (2010) and Pace and Zhu (2012)) have criticized the spatial lag model because it uses a single autocorrelation coefficient to characterize spatial dependence, or because it maintains the same autocorrelation spillover effect across all exogenous variables (and across the exogenous variables and the unobserved error term), we submit that these criticisms are themselves misguided because they miss out on the fundamental point that the basic specification of spatial dependence must not change whether a determining variable is in the observed portion or in the unobserved portion.

in the equations above by allowing a covariance in the error terms between the discrete and continuous dependent variables. Let the covariance matrix of the $[I \times 1]$ vector $\breve{\boldsymbol{y}}_q = (\breve{\boldsymbol{u}}_q, \eta_q)$ be specified as:

$$\mathrm{Cov}(\breve{\boldsymbol{y}}_q) = \begin{bmatrix} \breve{\boldsymbol{\Lambda}} & \boldsymbol{\Psi} \\ \boldsymbol{\Psi}' & \sigma^2 \end{bmatrix} \tag{5}$$

where $\boldsymbol{\Psi}$ is an $(I-1) \times 1$ vector capturing covariance effects between the $\breve{\boldsymbol{u}}_q$ vector and the η_q scalar (the level of covariance is assumed to be identical across households). All elements of the symmetric matrix above (of size $I \times I$) are identifiable after the scale normalization of the first element of $\breve{\boldsymbol{\Lambda}}$ (as discussed earlier). If the aspatial regression $y_q = \boldsymbol{\gamma}'\boldsymbol{z}_q + \eta_q$ were adopted, then, by specification, the covariance of $\breve{\boldsymbol{y}}_q$ is identically and independently distributed across observations. However, the situation changes dramatically as soon as the spatial structure of Eq. (3) is used. In particular, dependence across y_q values is generated by the spatial structure, and this further permeates in a secondary fashion into covariations in the utilities of alternatives of one household and the y_q measure of another household. To explicate this unique spatial dependency formulation, arrange the latent utility values of households and the error term vector in the continuous dependent variable across all households $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_Q)'$ into a $[Q(I-1) + Q] \times 1 = QI \times 1$ vector $\breve{\boldsymbol{u}}\boldsymbol{\eta} = (\breve{\boldsymbol{u}}', \boldsymbol{\eta}')'$. This vector is distributed with mean zero and a $QI \times QI$ covariance matrix given by:

$$\mathrm{Cov}(\breve{\boldsymbol{u}}\boldsymbol{\eta}) = \begin{bmatrix} (\mathbf{IDEN}_Q \otimes \breve{\boldsymbol{\Lambda}}) & (\mathbf{IDEN}_Q \otimes \boldsymbol{\Psi}) \\ (\mathbf{IDEN}_Q \otimes \boldsymbol{\Psi})' & (\mathbf{IDEN}_Q \otimes \sigma^2) \end{bmatrix}. \tag{6}$$

The covariance matrix above would also correspond to the distribution of the vector $(\breve{\boldsymbol{u}}\,\boldsymbol{y})'$ if there were no spatial dependence (because then the distribution of the vectors $\boldsymbol{y}$ and $\boldsymbol{\eta}$ are the same). However, the situation is different with the spatial dependence structure of Eq. (4). In this spatial dependence case, it may be shown that the $QI \times 1$ vector, $\breve{\boldsymbol{u}}\boldsymbol{y} = (\breve{\boldsymbol{u}}'\boldsymbol{y}')'$, is multivariate normally distributed with a $QI \times QI$ covariance matrix as follows:

$$\mathrm{Cov}(\breve{\boldsymbol{u}}\boldsymbol{y}) = \breve{\boldsymbol{\Omega}} = \begin{bmatrix} \mathbf{IDEN}_{Q(I-1)} & \mathbf{0}_{Q(I-1)\times Q} \\ \mathbf{0}_{Q\times Q(I-1)} & \boldsymbol{S} \end{bmatrix} \mathrm{Cov}(\breve{\boldsymbol{y}}\,\boldsymbol{\eta}) \begin{bmatrix} \mathbf{IDEN}_{Q(I-1)} & \mathbf{0}_{Q(I-1)\times Q} \\ \mathbf{0}_{Q\times(Q-1)} & \boldsymbol{S}' \end{bmatrix}, \tag{7}$$

where $\mathbf{0}_{Q(I-1)\times Q}$ is a zero matrix of dimension $Q(I-1) \times Q$.

The covariance matrix above corresponds to the vector $\breve{\boldsymbol{u}}\boldsymbol{y} = (\breve{\boldsymbol{u}}'\boldsymbol{y}')'$, where the vector $\breve{\boldsymbol{u}}$ represents the vectorization (across individuals) of the latent utility differentials taken with respect to the first alternative for each household. For estimation, however, what is needed is the covariance matrix for the vector $\boldsymbol{uy} = (\boldsymbol{u}'\boldsymbol{y}')'$, where the vector $\boldsymbol{u}$ represents the vectorization (across households) of the latent utility differentials taken with respect to the chosen alternative for each household. To compute this, first construct the general covariance matrix $\boldsymbol{\Omega}$ for the original $[Q(I+1)] \times 1$ vector $\boldsymbol{Uy} = (\boldsymbol{U}'\boldsymbol{y}')'$, while also ensuring all parameters are identifiable (note that $\boldsymbol{\Omega}$ is equivalently the covariance matrix of $\tau = (\boldsymbol{\varepsilon}', (\boldsymbol{S\eta})')'$). To do so, define a matrix $\boldsymbol{D}$ of size $[Q(I+1)] \times QI$. The first I rows and $(I-1)$ columns correspond to the first household. Insert an identity matrix of size $(I-1)$, after supplementing with a first row of zeros, in the first through Ith rows and the first through $(I-1)$ th columns of the matrix. The rest of the elements in the first I rows and the first $(I-1)$ columns take a value of zero. Next, rows $(I+1)$ through $2I$ and columns (I) through $2(I-1)$ correspond to the second household. Again position an identity matrix of size $(I-1)$ after supplementing with a first row of zeros into this position. Continue this for all Q households. Put zero values in all cells without any value up to this point. Finally, insert an identity matrix of size Q into the last Q rows and Q columns of the matrix $\boldsymbol{D}$. Thus, for the case with two households, if the nominal variable has 4 alternatives, the matrix $\boldsymbol{D}$ takes the form shown below:

$$\boldsymbol{D} = \begin{bmatrix} 0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \end{bmatrix}_{10\times 8} \tag{8}$$

Then, the general covariance matrix of $\boldsymbol{Uy}$ may be developed as $\boldsymbol{\Omega} = \mathbf{D}\breve{\boldsymbol{\Omega}}\mathbf{D}'$. All parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on utility differences and, at the same time, it provides a consistent means to obtain the covariance matrix $\widetilde{\boldsymbol{\Omega}}$ of $\boldsymbol{uy} = (\boldsymbol{u}'\boldsymbol{y}')'$ that is needed for estimation (and is with respect to each individual's chosen alternative for the nominal variable). Specifically, to develop the distribution for the vector $\boldsymbol{uy}$, define a matrix $\mathbf{M}$ of size $QI \times Q(I+1)$. The first $(I-1)$ rows and I columns correspond to the first household. Insert an identity matrix of size $(I-1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative of the first household. The rest of the columns for the first $(I-1)$ rows and the rest of the rows for the first I columns take a value of zero. Next, rows (I) through $2(I-1)$ and columns $(I+1)$ through $2I$ correspond to the second household. Again position an identity matrix of size $(I-1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative of the second household now. Continue this procedure for all Q households. Finally, insert an identity matrix of size Q into the last Q rows and Q columns of the matrix $\mathbf{M}$. With the matrix $\mathbf{M}$ as defined, the covariance matrix $\widetilde{\boldsymbol{\Omega}}$ is given by $\widetilde{\boldsymbol{\Omega}} = \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'$.

Next, define $\boldsymbol{\mu} = \mathbf{M}(\boldsymbol{V}|\boldsymbol{d})$, where the vector $\boldsymbol{V}$ is defined as earlier, $\boldsymbol{d} = \boldsymbol{Sz\gamma}'$ is a $(Q \times 1)$-vector, and $(\boldsymbol{V}|\boldsymbol{d})$ denotes the vertical concatenation of the vectors $\boldsymbol{V}$ and $\boldsymbol{d}$ so that $(\boldsymbol{V}|\boldsymbol{d})$ is a $(Q(I+1)\times 1)$ vector and $\boldsymbol{\mu}$ is a $(QI)\times 1)$ vector. Then, by construction, $\boldsymbol{uy} = (\boldsymbol{u}'\boldsymbol{y}')' \sim MVN_{QI}(\boldsymbol{\mu}, \widetilde{\boldsymbol{\Omega}})$. Partition $\boldsymbol{\mu}$ and $\widetilde{\boldsymbol{\Omega}}$ so that

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{c} \end{bmatrix}, \quad \text{and} \quad \widetilde{\boldsymbol{\Omega}} = = \begin{bmatrix} \boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_{uy} \\ \boldsymbol{\Sigma}'_{uy} & \boldsymbol{\Sigma}_y \end{bmatrix}, \tag{9}$$

where $\boldsymbol{g}$ is a $(Q(I-1)\times 1)$ subvector, $\boldsymbol{c}$ is a $(Q\times 1)$ subvector, $\boldsymbol{\Sigma}_u$ is a $Q(I-1)\times Q(I-1)$ submatrix, $\boldsymbol{\Sigma}_{uy}$ is a $Q(I-1)\times Q$ submatrix, and $\boldsymbol{\Sigma}_y$ is a $Q\times Q$ submatrix. The conditional distribution of $\boldsymbol{u}$, given $\boldsymbol{y}$, is multivariate normal with mean $\widetilde{\boldsymbol{g}} = \boldsymbol{g} + \boldsymbol{\Sigma}_{uy}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}-\boldsymbol{c})$ $[Q(I-1)\times 1\text{vector}]$ and variance $\widetilde{\boldsymbol{\Sigma}}_u = \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_{uy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}'_{uy}$ $[Q(I-1)\times Q(I-1)\text{matrix}]$.

Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\theta = [\boldsymbol{\beta}', \boldsymbol{\gamma}', \delta, (Vech(\breve{\boldsymbol{\Lambda}}))', (Vech(\boldsymbol{\Psi}))']'$, where $Vech(\breve{\boldsymbol{\Lambda}})$ represents the vector of upper triangle elements of $\breve{\boldsymbol{\Lambda}}$. Then the likelihood function may be written as:

$$L(\theta) = f_Q(\boldsymbol{y};\boldsymbol{c},\boldsymbol{\Sigma}_y) \times \Pr[(\boldsymbol{u}|\boldsymbol{y}) \leqslant 0], = \frac{1}{\prod_{q=1}^{Q}(\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y})_q}\phi_Q(\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y}^{-1}(\boldsymbol{y}-\boldsymbol{c});\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y}^{-1}\boldsymbol{\Sigma}_y\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y}^{-1}) \times \Phi_{(I-1)\times Q}[(\boldsymbol{\omega}_{\widetilde{\boldsymbol{\Sigma}}_u}^{-1}(-\widetilde{\boldsymbol{g}});\boldsymbol{\omega}_{\widetilde{\boldsymbol{\Sigma}}_u}^{-1}\widetilde{\boldsymbol{\Sigma}}_u\boldsymbol{\omega}_{\widetilde{\boldsymbol{\Sigma}}_u}^{-1}], \tag{10}$$

where $\phi_Q(.;.)$ is the multivariate normal density function of Q dimensions, $\Phi_{(I-1)\times Q}$ is the multivariate normal cumulative distribution function of $(I-1)\times Q$ dimensions, $\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y}$ is a diagonal matrix containing the square root of the diagonals of $\boldsymbol{\Sigma}_y$, $(\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y})_q$ represents the qth diagonal element of $\boldsymbol{\omega}_{\boldsymbol{\Sigma}_y}$, and $\boldsymbol{\omega}_{\widetilde{\boldsymbol{\Sigma}}_u}$ is a diagonal matrix containing the square root of the diagonals of $\widetilde{\boldsymbol{\Sigma}}_u$.

The above likelihood function involves the evaluation of an $(I-1)\times Q$ -dimensional integral, which is prohibitive even for medium-sized samples. So, the Maximum Approximate Composite Marginal Likelihood (MACML) approach of Bhat (2011), in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is used in this paper. Details of this MACML approach are available in Bhat (2011), but conceptually is based on replacing the second term in Eq. (10) with a surrogate function that uses the product of the joint probability of residential location choices of couplets of individuals. To ensure constraints on the autoregressive term δ (Eq. (2)), the analyst can parameterize $\delta = \pm 1/[1+\exp(\widetilde{\delta})]$. Once estimated, the $\widetilde{\delta}$ estimate can be translated back to an estimate of δ. If spatial dependency in the form of spillover or permeation effects exists, then a positive autoregressive parameter will be obtained.

4.4. Attributing VMT variation to different factors

The ultimate objective of this paper is to quantify the relative contributions of each of five factors to explaining variation in household VMT. The five factors are: (1) household and person socio-economic and demographic (SED) characteristics, (2) residential built environment (BE) attributes, (3) residential self-selection (SS) effects, (4) human socio-spatial dependency (SSD) effects, and (5) remaining unknown or omitted factors (UF). To accomplish this, start with Eq. (4),

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}\boldsymbol{\gamma} + \boldsymbol{S}\boldsymbol{\eta} \tag{11}$$

To determine an estimate of the SSD effect on the logarithm of VMT, compute the mean of the sum of squared regression of the above equation with the estimated δ value embedded in the $\boldsymbol{S}$ vector. This mean SSR (or the mean squared regression or MSR) includes all of the four (SED, BE, SS, and SSD) effects, with the mean sum of squared residuals (MSE) representing the effect of remaining unknown or unobserved factors (LNVMT_{UF}). Next, compute the mean MSR of the above equation with $\delta = 0$ (which is equivalent to an aspatial model). The difference between the two provides the variation in logarithm of VMT (at an average household level) explained by the SSD effect (label this $\text{LNVMT}_{\text{SSD}}$). Once the SSD effect is determined, consider the aspatial VMT regression at the household level (note that, after accommodating VMT error dependencies through the spatial-social dependence, engendered by the $\boldsymbol{S}$ matrix in Eq. (4), there is no remaining covariance in VMT across households):

$$y_q = \boldsymbol{\gamma}'\boldsymbol{z}_q + \eta_q \tag{12}$$

The vector $\boldsymbol{z}_q$ is now partitioned into variables that correspond to SED characteristics, and the two dummy BE variables that characterize the impact of two of the residential density alternatives (with the first residential density alternative serving as the base). Let $\boldsymbol{z}_q = (\boldsymbol{z}'_{q,SED}, \boldsymbol{z}'_{q,BE})'$, and correspondingly partition the $\boldsymbol{\gamma}$ vector into $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_{SED}, \boldsymbol{\gamma}'_{BE})'$. Also, for each individual, the error-differenced utilities for the second and third residential alternatives (in the choice model) are correlated with the corresponding η_q error term in the log(VMT) equation, based on Eq. (5). In the three alternative residential choice case, Eq. (5) may be rewritten as follows:

$$\text{Cov}(\widetilde{\boldsymbol{y}}_q) = Cov\begin{pmatrix} u_{q21} \\ u_{q31} \\ \eta_q \end{pmatrix} = Cov\begin{pmatrix} \xi_{q21} \\ \xi_{q31} \\ \eta_q \end{pmatrix} = \begin{bmatrix} 1 & \vartheta_{23} & \omega_{2\eta} \\ \vartheta_{23} & \vartheta_3^2 & \omega_{3\eta} \\ \omega_{2\eta} & \omega_{3\eta} & \sigma^2 \end{bmatrix}, \text{where} \tag{13}$$

$$\xi_{q21} = \varepsilon_{q2} - \varepsilon_{q1}, \xi_{q31} = \varepsilon_{q3} - \varepsilon_{q1}, \quad \breve{\boldsymbol{\Lambda}} = \begin{bmatrix} 1 & \vartheta_{23} \\ \vartheta_{23} & \vartheta_3^2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Psi} = \begin{bmatrix} \omega_{2\eta} \\ \omega_{3\eta} \end{bmatrix}.$$

Partition $\xi_{q21} = \tau_{2\eta} + \kappa_{q23} + \widetilde{\xi}_{q21}, \xi_{q31} = \tau_{3\eta} + \kappa_{q23} + \widetilde{\xi}_{q31}, \eta_q = \tau_{2\eta} + \tau_{3\eta} + \widetilde{\eta}_q$. Then, by construction, $\vartheta_{23} = Var(\kappa_{q23}), Var(\widetilde{\xi}_{q21}) = 1 - Var(\tau_{2\eta}) - Var(\kappa_{q23}), Var(\widetilde{\xi}_{q31}) = \vartheta_3^2 - Var(\tau_{3\eta}) - Var(\kappa_{q23})$, $\omega_{2\eta} = Var(\tau_{2\eta}), \omega_{3\eta} = Var(\tau_{3\eta})$, and $Var(\widetilde{\eta}_q) = \sigma^2 - \omega_{2\eta} - \omega_{3\eta}$.

With the notations above, rewrite Eq. (12) as:

$$y_q = \boldsymbol{\gamma}'_{SED}\boldsymbol{z}_{q,SED} + \boldsymbol{\gamma}'_{BE}\boldsymbol{z}_{q,BE} + \tau_{2\eta} + \tau_{3\eta} + \widetilde{\eta}_q. \tag{14}$$

The MSR of the above regression (label this as the aspatial SSR or ASSR) is exactly the same as the one computed earlier for Eq. (11) with $\delta = 0$ (there are no re-estimations of the model undertaken; the estimated coefficient values from the original model are all retained as such). This MSR can be split into that attributable to SED ($\text{LNVMT}_{\text{SED}}$) by computing the degradation in MSR after setting all elements of the $\boldsymbol{\gamma}_{SED}$ to zero. Next, keep the elements of the $\boldsymbol{\gamma}_{SED}$ vector at the estimated values, and set the elements of $\boldsymbol{\gamma}_{BE}$ to zero. The degradation in the MSR of this regression relative to ASSR can be attributable to BE (LNVMT_{BE}). Proceeding forward, the mean variation effect attributable to self-selection SS (LNVMT_{SS}) is essentially equivalent to $Var(\tau_{2\eta} + \tau_{3\eta}) = \omega_{2\eta} + \omega_{3\eta}$, which is obtained from the estimated covariances (see Eq. (13)). The variance of $\widetilde{\eta}_q$ is a measure of the unexplained variation in the aspatial model, but the overall unexplained variation is already captured in the spatial model as LNVMT_{UF}. Finally, to quantify the effect of each factor on VMT, simply compute the fraction of each exp(LNVMT) contribution as a proportion of the sum of the exponentials of VMTs from each contributing source (= $\exp(\text{LNVMT}_{\text{SED}}) + \exp(\text{LNVMT}_{\text{BE}}) + \exp(\text{LNVMT}_{\text{SS}}) + \exp(\text{LNVMT}_{\text{SSD}}) + \exp(\text{LNVMT}_{\text{UF}})$. These proportions provide the percentage contribution at the point median VMT estimate.

5. Model estimation results

This section presents a description of model estimation results. Many alternative model specifications were tested to arrive at the final model specification. It should be noted that a number of potential explanatory variables were not included in the residential location (density) choice utilities because of potential endogeneity effects. Variables such as dwelling unit type, vehicle ownership, and number of drivers may be regarded as endogenous to residential location choice and were hence omitted from the specification. Treating such choice variables as exogenous factors could lead to endogeneity bias and adversely affect inferences that may be drawn from the model estimation results. In addition, these dimensions are often closely related to density; for example, dense urban environments are likely to be characterized by multi-family residential dwelling units with high levels of transit service, thus reducing the need for owning cars or possessing a driver's license. Including these dimensions as exogenous variables would render it difficult to isolate the self-selection effects from exogenous variable effects because these variables are actually part of the self-selection phenomenon. It should be recognized that the omission of these variables from the residential location density model may affect inferences drawn because at least some of the effects of these variables is likely to be subsumed in the unexplained portion of the variance in household VMT. The development of a more complex multi-dimensional simultaneous equations model system that can jointly model these varied choices in an integrated framework, and thus account for their effects appropriately, remains a task for future research. It should, however, be noted that a number of such variables are included in the household VMT regression equation.

Another key consideration in the specification of the model system estimated in this study is that no additional built environment attributes are introduced in the VMT equation to avoid an entanglement of built environment attributes embedded in the residential choice definition with any other built environment attributes that could be introduced separately in the household VMT regression equation. This remains a methodological challenge to be addressed in future research efforts. In the model system estimated in this study, the residential location choice alternatives are defined by density; and density variables are, in turn, incorporated in the VMT equation to capture built environment effects. By doing so, it is possible to explicitly and easily tease out residential sorting effects (represented by error covariances) from built environment effects. If additional built environment attributes were included in the VMT equation specification, the correlation between density and these additional built environment attributes would render it difficult to cleanly separate residential self-selection effects from true built environment effects. Future research efforts should focus on the development of a model formulation where it is possible to separately measure residential self-selection effects and true built environment effects in the presence of multi-collinearity between the residential location choice descriptor(s) and the built environment attributes.

Repeated attempts were made to estimate a full model specification with spatial dependency. A variety of spatial dependency forms were specified (including two different versions of the contiguity-based definition, inverse distance and its power functions, the inverse of exponential distance, and the shared edge length between zones) and used to define the weight matrices that represent strength of association between observations. Every specification that was attempted yielded a spatial dependency or autoregressive parameter (δ) that was not statistically significantly different from zero. Reasons for the statistical insignificance of the spatial dependency parameter are not immediately clear, but the fact that the parameter was repeatedly found to be insignificant for a large variety of specifications suggests that the spatial dependency effects may truly be insignificant in this particular data set, or there are other unknown forces at play that are rendering this effect to be non-existent. One possible explanation for this is that significant peer or neighbor effects may not exist for household VMT (after controlling for socio-demographic and built environment attributes). Household VMT may largely be determined by intra-household interactions and task allocation among household members. Social and spatial diffusion effects arising from interactions among neighbors may not be playing an important role in shaping household-level VMT. Due to the insignificance of the spatial dependency effect, only the final non-spatial or aspatial model estimation results are presented. In addition, the allocation of household VMT to various contributing factors omits spatial dependency effects and only considers the three other effects (socio-economic and demographic, residential self-selection, and built environment) together with unexplained or unknown effects.

Model estimation results for the aspatial model with self-selection are shown in Table 2. An independent model that ignores self-selection effects engendered through error covariances was also estimated; results for that model are quite similar to those seen in the

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)	Natural Log of 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 variables [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 household 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)
Number of vehicles in household				
One vehicle	*	–	–	1.6606 (22.35)
Two or more vehicles	*	–	–	2.5955 (32.45)
Number of workers in household	*	–	–	0.1505 (4.70)
Presence of students in household	*	–	–	0.1388 (2.55)
Fraction of unemployed in household	*	–	−0.3073 (−3.54)	–

"*" Denotes that the category considered is the base alternative.
"–" Denotes that the variable is statistically insignificant and so is not included in the specification.

model with self-selection and hence the table of results for the independent model system is omitted. The residential location (density) model component takes the form of a multinomial probit (MNP) model while the vehicle miles of travel model component takes the form of a continuous log-linear regression model.

In the MNP model of residential location (density) choice, it can be seen that alternative specific constants for the medium and high density categories are negative, suggesting that *ceteris paribus*, households are more likely to locate in low density neighborhoods. Single persons are more likely, however, to locate in high density neighborhoods. Consistent with descriptive statistics seen earlier and prior research (Cao and Fan, 2012), lower income households are more likely to locate in medium- and higher density neighborhoods, as are households belonging to ethnic minority segments (African-American and Hispanic). Households with a higher fraction of unemployed individuals are less likely to locate in high density neighborhoods, presumably because households in low density neighborhoods are of larger sizes and with children (who are naturally unemployed).

In the continuous linear regression model, the fraction of individuals in the household in the middle age groups is positively associated with household VMT production, presumably because such households are at a lifecycle stage that is associated with a high level of trip-making, compared to households with a higher fraction of individuals in older age groups (Collia et al., 2003). Residential location (density) is found to significantly affect household VMT, consistent with the pattern seen in Fig. 1 and as reported extensively in the literature. Households in medium and high density neighborhoods produce fewer VMT as evidenced by the negative coefficients, with the effect amplified in the context of high density areas relative to medium density areas. As expected, vehicle ownership is a strong predictor of household VMT with multi-vehicle owning households likely to generate more VMT than other vehicle ownership groups.

A review of error variance-covariance estimates in the matrix $\breve{\boldsymbol{\Lambda}}$ for the independent model system (where error covariances across the discrete choice and linear regression model components are restricted to zero; that is, all elements of the matrix $\boldsymbol{\Psi}$ are set to zero) and the joint model system (that accounts for self-selection effects through the elements of the $\boldsymbol{\Psi}$ matrix) reveals a statistically significant covariance between density categories in the residential location choice model. In particular, referring to Eq. (13), the estimated value of ϑ_{23} is 0.4437 (*t*-statistic of 8.12), and that of ϑ_3^2 is 1.002 (*t*-statistic of 9.60) in the joint system (these estimated values were similar in the independent model). The positive value of ϑ_{23} suggests that unobserved attributes that contribute to living in a medium (high) density configuration positively contribute to residing in a high (medium) density area (though, very technically, the matrix $\breve{\boldsymbol{\Lambda}}$ is a differenced utility matrix with respect to the low density category). This result is consistent with expectations. Attitudes and lifestyle preferences that motivate an individual to seek residential locations in higher density areas are likely to positively influence choice of residence in both medium and high density neighborhoods.

In the model with self-selection, it is found that significant error covariances exist between residing in medium or high density neighborhoods (relative to low density living) and vehicle miles of travel. Specifically, the estimated values of $\omega_{2\eta}$ and $\omega_{3\eta}$ are 0.108 (*t*-statistic of 2.19) and 0.089 (*t*-statistic of 1.92), respectively. These significant error covariances demonstrate the importance of modeling these choices (i.e., residential location and household VMT) jointly in a simultaneous equations modeling framework capable of accounting for shared unobserved attributes affecting multiple endogenous variables of interest. What is interesting is that both error covariances are positive and significant. In other words, unobserved attributes that contribute to residing in higher density neighborhoods (relative to residing in low density neighborhoods) also contribute to an increase in household VMT after accounting for observed exogenous covariates included in the model specification. Although this may appear counter-intuitive at first glance, it is not necessarily so. The very unobserved attributes that contribute to seeking residential location in higher density neighborhoods may very well contribute to higher VMT production. After controlling for built environment attributes and household socio-economic and demographic characteristics, households that favor active lifestyles and seek a variety of activity opportunities (latent unobserved traits) are likely to undertake more travel and hence produce more VMT than observationally equivalent households that have different (more sedentary) lifestyle preferences.

An examination of goodness-of-fit statistics reveals that the composite log-likelihood value for the joint model with 25 parameters is −10,233.30 while the corresponding value for the independent model (with 23 parameters) is −10,236.39. The goodness-of-fit of the two models may be compared using the adjusted composite likelihood ratio test (ADCLRT) statistic that is approximately χ^2 distributed (Bhat, 2011). The ADCLRT statistic value is 6.04, which is larger than the critical χ^2 table value with two degrees of freedom at a 95% confidence level. This shows that the model with self-selection offers a statistically significant, but not necessarily very large, improvement in fit to the data.

6. Discussion and conclusions on the relative contribution of factors to household VMT

Because the spatial dependency parameter was found to be statistically insignificant across a wide range of specifications, model results were used to apportion the contribution of three factors to explaining the variance in household VMT. These include: (1) household socio-economic and demographic characteristics (SED); (2) built environment attributes of the residential zone (BE); and (3) residential self-selection effects (SS). Any unexplained portion of the variance in VMT may then be attributed to unknown unobserved attributes or variables omitted in the specification. It is possible that there are spatial dependency effects that the tested model specifications were not able to capture; if such effects truly do exist, then they would be absorbed into the unexplained portion as well.

We next adopt the methodology described in Section 4 (albeit with a slight simplification to account for the lack of significant spatial dependency effects). The results reveal that socio-economic and demographic characteristics explain 33 percent of the variation in household VMT. Built environment attributes, after controlling for self-selection effects, explain 12 percent of the variation in VMT; self-selection effects (captured through error covariances) account for 11 percent of the variation in household VMT. That leaves 44 percent of the variance in household VMT unexplained by socio-economic characteristics, residential built environment attributes, and self-selection effects considered in the model specifications of this paper. From a regression analogy, that is akin to achieving an R^2 goodness-of-fit value of 0.56, which is quite consistent with (and even better than) typical goodness-of-fit statistics obtained when estimating household-level regression models of trip-making. Within the 56 percent of household VMT variance that is explained by the three factors, the results suggest that socio-economic and demographic characteristics account for 58.5 percent of the explained portion, residential self-selection accounts for 19.8 percent, and built environment attributes account for 21.7 percent. Focusing on the land use (built environment) and self-selection effects, it appears that 48 percent of the built environment effect is attributable to self-selection, leaving the remaining 52 percent as the *true* built environment effect.

The relative contribution of various effects to explaining household VMT found in this paper suggests that household socio-economic and demographic characteristics play a significant and large role (much larger than built environment and self-selection) in shaping household VMT, a finding that has been reported by others (Badoe and Miller, 2000). Most previous studies that study the contribution of various factors to explaining VMT do so in the context of separating true built environment effects from residential self-selection effects. As noted earlier, Bhat et al. (2014) and Cao and Fan (2012) found that the self-selection effect is considerably less than the built environment effect. This study finds a more even split between self-selection and true built environment effects, suggesting that the relative contribution of these two effects may vary across geographic contexts. It should be noted that this study utilized a New York region data set, while the Cao and Fan (2012) study utilized a data set from North Carolina and Bhat et al. (2014) used a data set from San Francisco. Another important consideration is that this paper examines household VMT as the dependent variable of interest; the Cao and Fan (2012) paper examines self-selection effects in the context of person miles of travel. It is plausible that the relative contribution of these effects varies based on the choice of dependent variable. Indeed, Cao and Fan (2012) find that the contribution of self-selection effects can be as high as 64% and 49% for driving duration and transit duration, respectively, while Bhat et al. (2014) report self-selection contributions of the order of 41 percent and 45 percent for the number of non-motorized and motorized tours, respectively. In addition, it is possible that the absence of a rich set of built environment attributes in the household

VMT generation equation may have led to an under-estimation of the true built environment effects in this study. However, given that density is generally highly correlated with other built environment attributes, it is uncertain whether the inclusion of additional built environment variables in the model specification would necessarily yield a (much) greater built environment effect than that obtained here (in the context of explaining household VMT).

There are two noteworthy aspects revealed in the analysis of this paper. First, spatial dependency effects may not be all that significant in explaining household VMT. In the case of person-level VMT, it is likely that dependency effects play a larger role because persons interact, at a minimum, with other household members. Nevertheless, given the large body of literature that has found significant spatial dependency effects in the context of modeling activity-travel choices (Adjemian et al., 2010; Paleti et al., 2013a), this is worthy of additional investigation. Second, this paper finds that household socio-economic and demographic variables play a much larger role in explaining household VMT variation than built environment and residential self-selection effects (combined). This is not surprising, given that household VMT is naturally dependent on household structure, income, and size. An examination of person-level VMT (where VMT is scaled to a per-capita basis) may offer additional insights on the relative contribution of socio-economic and demographic characteristics vis-à-vis other built environment and self-selection effects. The use of household VMT as the dependent variable in this study may have resulted in an amplified estimate of the relative contribution of household socio-economic and demographic characteristics (because larger households will likely generate greater VMT). Nonetheless, the study results suggest that changes in built environment attributes (as considered in the model specification) may not necessarily bring about substantial shifts in household-level VMT, possibly due to many other factors that remain unknown or unmeasured in typical household travel surveys. Indeed, recent evidence (Polzin, 2016) suggests that vehicle miles of travel (both in terms of aggregate total and on a per-capita basis) is attaining record high levels in the United States with the recovery of the economy from the deep recession, despite a number of trends (e.g., gentrification and transit-oriented urban development, millennials less auto-oriented than prior generations, and an increase in environmental consciousness) that would have otherwise reduced VMT generation. Overall, based on the results from our analysis and earlier literature, there is a suggestion that socio-economic effects drive much of the VMT changes as opposed to land use and neighborhood location self-selection effects.

As with any modeling study, the results obtained here are sensitive to the model specification. The percent of explained or unexplained variance, and the relative magnitudes of the effects of different factors in explaining household VMT, is naturally dependent on the model specification. The data set used in this study is derived from a typical regional household travel survey and zonal land use information. The analysis is naturally limited by the set of variables and built environment attributes available in the data, and the results are sensitive to the final model specification adopted in the study. Another limitation that should be noted is that zonal density is used as the sole descriptor of the built environment; although density is strongly correlated with other built environment attributes, it is possible that the inclusion of additional built environment measures as explanatory variables in the household VMT equation could have yielded a larger true built environment effect than that reported here. Future research efforts should focus on including additional built environment attributes (such as proximity to transit infrastructure and land use diversity) in the model specification to more accurately quantify the relative contribution of built environment variables in explaining household VMT. However, this also calls for the development of a methodological formulation capable of unraveling true built environment effects from residential self-selection effects when explanatory variables correlated with the residential location descriptor variable(s) are included in the household VMT equation. Another area for future research involves the further exploration of social-spatial dependency effects that were found to be insignificant in this study. Future research efforts should strive to employ a finer spatial resolution with a view to better capture interaction effects that could be masked when using more aggregate zonal definitions. Finally, there is a need for fundamental research that aims to advance an understanding of the drivers of household VMT. In this study, it is found that 44% of the variance remains unexplained, clearly pointing to the limited understanding in the profession of the factors that shape household VMT. Research efforts aimed at identifying causal factors that contribute to household and person VMT should continue, with a view to help inform land use design and transportation policy.

Acknowledgements

The authors gratefully acknowledge the comments provided by three anonymous reviewers that helped improve the paper. This research was partially supported by the Center for Teaching Old Models New Tricks (TOMNET) as well as the Data-Supported Transportation Operations and Planning (D-STOP) Center, both of which are Tier 1 University Transportation Centers sponsored by the U.S. Department of Transportation. The authors are grateful to Lisa Macias for her help in formatting this document. The authors are solely responsible for the contents of the paper.

References

AASHTO, 2013. Commuting in America 2013 – The National Report on Commuting Patterns and Trends: The Role of Commuting in Overall Travel. < http://traveltrends.transportation.org/Documents/B2_CIA_Role%20Overall%20Travel_web_2.pdf > (accessed July 27, 2017).

Adjemian, M.K., Lin, C.Y.C., Williams, J., 2010. Estimating spatial interdependence in automobile type choice with survey data. Transport. Res. Part A 44 (9), 661–675.

Anselin, L., 2003. Spatial externalities, spatial multipliers and spatial econometrics. Int. Reg. Sci. Rev. 26, 153–166.

Badger, E., 2014. The Many Reasons Millennials are Shunning Cars. Washington Post. < https://www.washingtonpost.com/news/wonk/wp/2014/10/14/the-many-reasons-millennials-are-shunning-cars/?utm_term=.f2ef038628b6 > (accessed July 25, 2017).

Badoe, D.A., Miller, E.J., 2000. Transportation–land-use interaction: empirical findings in North America, and their implications for modeling. Transport. Res. Part D 5 (4), 235–263.

Bastian, A., Börjesson, M., Eliasson, J., 2016. Explaining "peak car" with economic variables. Transport. Res. Part A 88, 236–250.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transport. Res. Part B 45 (7), 923–939.
Bhat, C.R., 2015. A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. Transportation 42 (5), 879–914.
Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transport. Res. Part B 41 (5), 506–526.
Bhat, C.R., Astroza, S., Hamdi, A., 2017. A spatial generalized ordered-response model with skew normal kernel error terms with an application to bicycling frequency. Transport. Res. Part B 95, 126–148.
Bhat, C.R., Astroza, S., Bhat, A.C., Nagel, K., 2016. Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: application to residential self-selection effects analysis in an activity time-use behavior model. Transport. Res. Part B 91, 52–76.
Bhat, C.R., Astroza, S., Sidharthan, R., Alam, M.J.B., Khushefati, W.H., 2014. A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. Transport. Res. Part B 68, 31–51.
Bhat, C.R., Sener, I.N., Eluru, N., 2010. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. Transport. Res. Part B 44 (8–9), 903–921.
Brownstone, D., Golob, T.F., 2009. The impact of residential density on vehicle usage and energy consumption. J. Urban Econ. 65 (1), 91–98.
Cao, X., 2010. Exploring causal effects of neighborhood type on walking behavior using stratification on the propensity score. Environ. Plan. A 42 (2), 487–504.
Cao, X., Fan, Y., 2012. Exploring the influences of density on travel behavior using propensity score matching. Environ. Plan. B 39 (3), 459–470.
Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: density, diversity, and design. Transport. Res. Part D 2 (3), 199–219.
Cervero, R., Murakami, J., 2010. Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. Environ. Plan. A 42 (2), 400–418.
Collia, D.V., Sharp, J., Giesbrecht, L., 2003. The 2001 National Household Travel Survey: a look into the travel patterns of older Americans. J. Saf. Res. 34 (4), 461–470.
Copperman, R., Bhat, C.R., 2007. An analysis of the determinants of children's weekend physical activity participation. Transportation 34 (1), 67–87.
Dill, J., Voros, K., 2007. Factors sffecting bicycling demand. Transport. Res. Rec.: J. Transport. Res. Board 2031, 9–17.
Economic Research, 2017. Moving 12-Month Total Vehicle Miles Traveled. < https://fred.stlouisfed.org/series/M12MTVUSM227NFWA > (accessed July 27, 2017).
Elhorst, P., Vega, S.H., 2013. On Spatial Econometric Models, Spillover Effects, and W. ERSA Conference Papers, ersa13p222, European Regional Science Association.
Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. J. Am. Plan. Assoc. 76 (3), 265–294.
Ewing, R., Hamidi, S., Grace, J.B., 2016. Compact development and VMT-environmental determinism, self-selection, or some of both? Environ. Plan. B 43 (4), 737–755.
Ferdous, N., Pendyala, R.M., Bhat, C.R., Konduri, K.C., 2011. Modeling the influence of family, social context, and spatial proximity on use of nonmotorized transport mode. Transport. Res. Rec. 2230, 111–120.
Frank, L.D., Engelke, P.O., 2001. The built environment and human activity patterns: exploring the impacts of urban form on public health. J. Plan. Lit. 16 (2), 202–218.
Handy, S., Cao, X., Mokhtarian, P.L., 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transport. Res. Part D 10 (6), 427–444.
Handy, S., Cao, X., Mokhtarian, P.L., 2006. Self-selection in the relationship between the built environment and walking: empirical evidence from Northern California. J. Am. Plan. Assoc. 72 (1), 55–74.
Heinen, E., Van Wee, B., Maat, K., 2010. Commuting by bicycle: an overview of the literature. Transp. Rev. 30 (1), 59–96.
Keane, M., 1992. A note on identification in the multinomial probit model. J. Bus. Econ. Stat. 10 (2), 193–200.
Kim, J., Brownstone, D., 2013. The impact of residential density on vehicle usage and fuel consumption: evidence from national samples. Energy Econ. 40, 196–206.
Krizek, K.J., 2003. Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? J. Am. Plan. Assoc. 69 (3), 265–281.
Lee, C., Moudon, A.V., 2006. The 3Ds + R: quantifying land use and urban form correlates of walking. Transport. Res. Part D 11 (3), 204–215.
Levy, J.I., Buonocore, J.J., Von Stackelberg, K., 2010. Evaluation of the public health impacts of traffic congestion: a health risk assessment. Environ. Health 9 (1), 65–77.
Logan, G., 2014. Millennials' Influence on New Home Demand. RCLCO. < http://www.rclco.com/advisory-millennials-real-estate > (accessed July 25, 2017).
McDonald, N.C., 2007. Travel and the social environment: evidence from Alameda County, California. Transport. Res. Part D 12 (1), 53–63.
Millard-Ball, A., Schipper, L., 2011. Are we reaching peak travel? Trends in passenger transport in eight industrialized countries. Transp. Rev. 31 (3), 357–378.
Munkin, M.K., Trivedi, P.K., 2008. Bayesian analysis of the ordered probit model with endogenous selection. J. Econ. 143 (2), 334–348.
New York Metropolitan Transportation Council, 2011. 2010/2011 Regional Household Travel Survey (RHTS). < https://www.nymtc.org/DATA-AND-MODELING/Travel-Surveys/2010-11-Travel-Survey > (accessed July 28, 2017).
Pace, R.K., Zhu, S., 2012. Separable spatial modeling of spillovers and disturbances. J. Geogr. Syst. 14 (1), 75–90.
Paleti, R., Bhat, C.R., Pendyala, R.M., Goulias, K.G., 2013a. Modeling of household vehicle type choice accommodating spatial dependence effects. Transport. Res. Rec.: J. Transport. Res. Board 2343, 86–94.
Paleti, R., Bhat, C.R., Pendyala, R.M., 2013b. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Transport. Res. Rec.: J. Transport. Res. Board 2382, 162–172.
Pinjari, A.R., Bhat, C.R., Hensher, D.A., 2009. Residential self-selection effects in an activity time-use behavior model. Transport. Res. Part B 43 (7), 729–748.
Pinkse, J., Slade, M.A., 2010. The future of spatial econometrics. J. Reg. Sci. 50 (1), 103–117.
Polzin, S., 2016. So Much for Peak VMT. Planetizen. < https://www.planetizen.com/node/84877/so-much-peak-vmt > (accessed July 27, 2017).
Sacramento Area Council of Governments, 2016. 2016 Metropolitan Transportation Plan/Sustainable Communities Strategy: Chapter 5B Vehicle Miles Traveled and Roadway Congestion Trends and Performance. < http://www.sacog.org/sites/main/files/file-attachments/chapter_5b_vehicle_miles_traveled_and_roadway_congestion_trends_and_performance.pdf > (accessed July 26, 2017).
Salvy, S.J., Roemmich, J.N., Bowker, J.C., Romero, N.D., Stadler, P.J., Epstein, L.H., 2009. Effect of peers and friends on youth physical activity and motivation to be physically active. J. Ped. Psychol. 34 (2), 217–225.
Sener, I.N., Bhat, C.R., 2012. Flexible spatial dependence structures for unordered multinomial choice models: formulation and application to teenagers' activity participation. Transportation 39 (3), 657–683.
You, D., Garikapati, V.M., Pendyala, R.M., Bhat, C.R., Dubey, S., Jeon, K., Livshits, V., 2014. Development of vehicle fleet composition model system for implementation in activity-based travel model. Transport. Res. Rec.: J. Transport. Res. Board 2430, 145–154.
Zhang, L., Hong, J., Nasri, A., Shen, Q., 2012. How built environment affects travel behavior: a comparative analysis of the connections between land use and vehicle miles traveled in US cities. J. Transp. Land Use 5 (3), 40–52.
Zhou, B., Kockelman, K., 2008. Self-selection in home choice: use of treatment effects in evaluating relationship between built environment and travel behavior. Transport. Res. Rec.: J. Transport. Res. Board 2077, 54–61.