Vehicle miles traveled and the built environment: evidence from vehicle safety inspection data

IRES2014-016

IRES Working Paper Series

Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data

Mi Diao, Joseph Ferreira, Jr.

May 2014

Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data

(*Forthcoming in Environment and Planning A*)

Mi Diao
Assistant Professor
Department of Real Estate
National University of Singapore
4 Architecture Drive
Singapore 117566
rstdm@nus.edu.sg

Joseph Ferreira, Jr.
Professor of Urban Planning and Operations Research
Department of Urban Studies and Planning
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139, USA
jf@mit.edu

Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data

Abstract:

This study examines the linkage between household vehicle usage and their residential locations within a metropolitan area using a newly available administrative dataset of annual private passenger vehicle safety inspection records (with odometer readings) and spatially detailed data on the built environment. Vehicle miles travelled (VMT) and a set of comprehensive built-environment measures are computed for a statewide 250m*250m grid cell layer using advanced Geographic Information Systems and database management tools. We apply factor analysis to construct five factors that differentiate the built-environment characteristics of the grid cells and then integrate the built-environment factors into spatial regression models of household vehicle usage that account for built environment, demographics, and spatial interactions. The empirical results suggest that built-environment factors not only play an important role in explaining the intra-urban variation of household vehicle usage, but may also be underestimated by previous studies that use more aggregate built-environment measures. One standard deviation variations in the built-environment factors are associated with as much as 5,000 mile differences in annual VMT per-household. This study also demonstrates the potential value of new georeferenced administrative datasets in developing indicators that can assist urban planning and urban management.

Key words:

Vehicle Miles Travelled, Built Environment, Vehicle Safety Inspection Data, GIS

1. INTRODUCTION

In the last few decades, the rapid growth of Greenhouse Gas (GHG) concentration in the atmosphere and associated negative effects of global warming are raising concerns worldwide. Policy makers are taking increasing steps to reduce GHG emissions and promote sustainable growth. The transportation sector is currently responsible for one quarter of the world's energy-related GHG emissions (Price et al. 2006), and personal mobility consumes two thirds of the total transportation energy use (IEA 2004). Improving vehicle and fuel technologies can mitigate some negative effects of driving, but additional steps are desirable and authorities and researchers have increasingly focused on the potential contribution of land use strategies in transportation GHG emission reduction, particularly those characterized by "smart growth". The design and implementation of such strategies require a good understanding of the linkage between vehicle usage of households and their residential locations.

The relationship between transportation and the built environment has long been studied and is recognized as complex. Based on the consumer choice theory, a number of researchers argue that the built environment can influence travel behavior through its differentiated impact on modal supply in terms of quantity and quality or latent effects, which then influence the consumption of travel in the short run. (e.g., Boarnet and Crane 2001; Zhang 2004). In the long run, the built environment can influence the location choices of households, and the consequent travel decisions (Handy et al. 2005). In addition, the built environment could also affect travel behavior in an indirect way through its impact on attitudes over time (Handy et al. 2005). Detailed reviews of related research can be found in Handy (1996), Boarnet and Crane (2001), and Ewing and Cervero (2001).

Household or individual-based survey data are the preferred instrument for empirical analysis of travel behavior, because the unit of analysis, an individual, can be readily associated with mode availability, travel cost, demographic factors, and built-environment measures. However, the high expense of individual travel surveys tends to limit the sample size[1] and update frequency, and privacy concerns often limit the geographic specificity with which trip origins and destinations can be revealed. These issues constrain the capability of survey-based studies in providing confidence in statistical accuracy at the neighborhood level.

With the development of spatial information infrastructure in the past few decades, the amount of location-tagged administrative data has been rapidly increasing, such as records for vehicle safety inspections, housing transactions, and transit fare card transactions. Although these datasets are originally designed to support narrow functions, they provide researchers and planning agencies an alternative data source to traditional travel surveys with lower collection cost, broader temporal and spatial coverage, and higher update frequency. These data could have wide-ranging applications in metropolitan planning and urban management.

This research examines the potential value as well as the difficulties of using georeferenced administrative data to construct performance measures and calibrate urban models to support government efforts in reducing transportation GHG emissions and promoting sustainable metropolitan growth. We take advantage of a newly-available administrative dataset, the odometer readings from annual safety inspections for all private passenger vehicles registered in Metro Boston to examine the relationship between Vehicle Miles Traveled (VMT) and various built-environment and demographic characteristics at a 250mx250m grid cell level.

[1] The sample size required to detect a statistically significant effect does not depend on the size of the population. However, the typical travel survey is not sufficiently large to test and refine models that differentiate spatial effects in great detail.

1.1 Methodological Background

A similar modeling approach with aggregate data has been adopted by other researchers. For example, Miller and Ibrahim (1998) carry out an empirical investigation into the relationship between the built environment and automobile travel at traffic analysis zone (TAZ) level in the Greater Toronto Area. Yang (2008) finds that urban spatial structure alone could explain a significant portion of the intra-region variation in commuting distance at the census tract level. Lindsey et al. (2011) find significant association between VMT and urban form characteristics at a 10-square-mile grid cell level.

These aggregate approaches[2] do not allow for an exploration of underlying factors and mechanisms by which the built environment may influence individual decisions. Despite these ecological fallacy risks, such studies do reveal underlying general spatial patterns of mobility within metropolitan areas and they provide useful insights for promoting urban growth in a more sustainable fashion. By using more disaggregated data and improved spatial modeling techniques, our approach can address some of the shortcomings of the previous studies:

(1) The zones used in previous aggregate studies are usually quite large. They use traffic analysis zones (TAZ), zip codes, or entire cities. At such an aggregated level, the intra-zone variations of the built-environment and demographic measures could be too large to ignore.

(2) The impact of the built environment on travel behavior is not constrained by the boundary of the neighborhood that a household lives in. The potential spatial autocorrelation may cause significant biases in previous aggregate studies.

[2] In aggregate analysis, both the build environment and travel behavior are measured at an aggregate level, such as cities, neighborhoods or zones. By contrast, disaggregate analysis rely on micro-level characteristics to investigate individuals' or households' travel behavior.

(3) Most previous studies rely on data aggregated from travel surveys. Therefore they have shortcomings similar to the survey-based disaggregate studies, such as high data collection expenses, low update frequencies, etc.

This study contributes to the literature of land use and transportation studies in multiple ways. Our analysis is conducted at the 250×250m grid cell level, which is small enough to support fine-grained characterization of the built environment. To provide a better characterization of the built environment, we compute a comprehensive set of built-environment indicators, and apply factor analysis to address multicollinearity. We calibrate spatial regression models to control for the potential spatial autocorrelation and explore the built-environment effects in different types of neighborhoods. Furthermore, the VMT dataset that we use is a good example of voluminous and repeatable administrative datasets that are becoming more potentially relevant to urban and transportation planning, because they can now be standardized and cross-referenced in useful ways with other datasets.

This paper is structured as follows: Section 2 introduces the study area and the major datasets. Section 3 describes the analysis techniques. Section 4 presents and discusses results of the empirical study. Section 5 summarizes the research findings and proposes future studies.

2. STUDY AREA, DATA, AND SPATIAL UNIT OF ANALYSIS

We select the Boston Metropolitan Area as the study area. Greater Boston (Figure 1) provides a large metro area that exhibits a broad range of built-environment characteristics[3] while remaining predominately within the one state for which disaggregated VMT and built-environment characteristics are available.

[3] Massachusetts Cities and Towns in Figure 1 are shaded based on their community type as designated by the Metropolitan Area Planning Commission (MAPC).

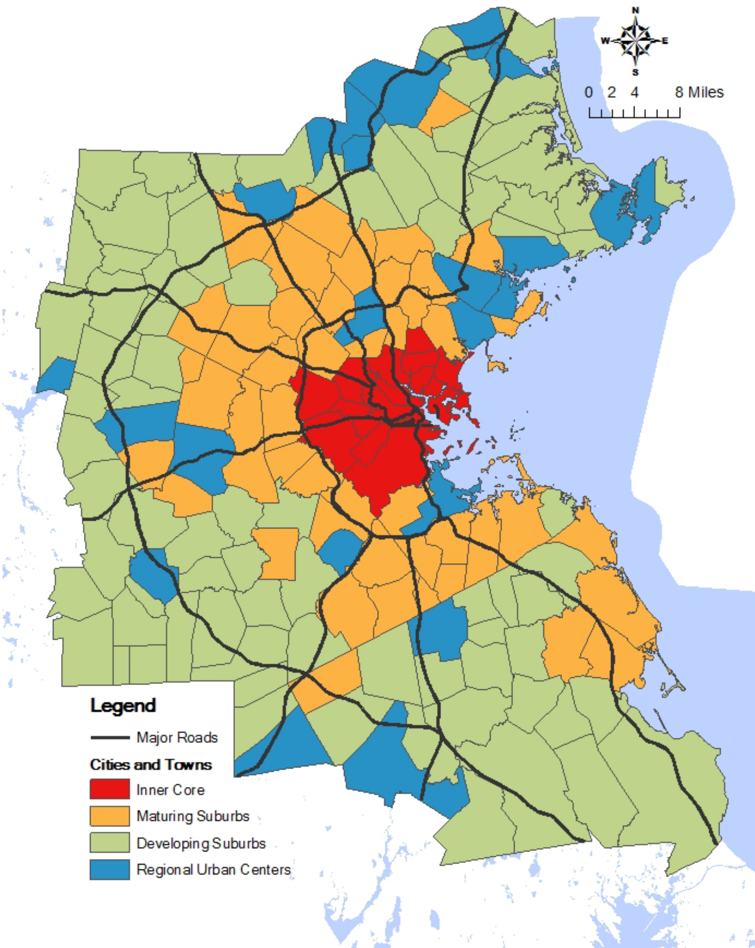


Figure 1: Metro Boston Grouped by Community Types

Safety inspection is mandated annually beginning within one week of registering a new or used vehicle in Massachusetts. The inspection utilizes computing equipment that records vehicle identification number (VIN) and odometer reading and transmits these data electronically to the Registry of Motor Vehicles (RMV) where they can be associated with the street address of the place of residence of the vehicle owner. In another study, Ferreira et al. (2013) uses the same dataset to estimate the VMT implications of alternative metropolitan growth scenarios in Metro Boston, but they do not develop built-environment measures or estimate regression models that relate VMT to built-environment and demographic factors. Holtzclaw (1994) and Holtzclaw et al. (2002) use similar datasets, odometer readings from auto emission inspections (smog checks) to explore the relationship between vehicle ownership and usage and urban-design characteristics at TAZ and zip codes level and find significant correlation. But new vehicles are exempted from smog checks for the first two years. Therefore their measure systematically biases VMT downwards for zones with large numbers of new vehicles (Brownstone 2008).

The vehicle safety inspection data have several advantages compared to the traditional travel surveys, thus providing an alternative perspective to examine the land use and transportation dynamics in a metropolitan area.

Spatial and temporal coverage: The dataset allows for studying the usage of millions of vehicles in the metropolitan area over a longer time period, compared to a few thousand individuals' movements within 1-2 days usually collected through travel surveys. Admittedly, well-designed surveys could provide good inference of the population with relatively small samples. However there are not many respondents included in any one neighborhood, which limits the efforts to adequately understand travel patterns for small areas (Handy, 1996).

Accuracy: Unlike travel surveys, this dataset does not depend on the subjects' willingness or ability to remember and report their driving. The Energy Information Administration (EIA)'s 1994 Residential Transportation Energy Consumption Survey shows that self-reported VMT values are 13% greater than odometer-based VMT in urban areas. EIA suggests that odometer-based VMT should be obtained if possible (Schipper and Moorhead 2000).

Update frequency: the dataset is continuously updated, which could lead to more reliable and trackable urban performance indicators that could support more prompt policy responses to emerging urban issues.

Collection cost: the odometer readings are routinely collected by the RMV as part of the annual vehicle safety inspection. While there are further data preparation and analysis costs to geocode and construct the analytic datasets, the bulk of the data collection costs are already covered by the $29 annual safety inspection fee.

The vehicle safety inspection data do have some drawbacks: (1) socio-economic and demographic attributes are not available due to privacy concerns. Such data are indispensable to calibrate models at a disaggregate level to explore the underlying behavior mechanism of individual mobility choice; and (2) the dataset is not primarily designed for modeling purposes and is not in an easy-to-use format. The gap between what the raw data directly offer and what is needed for urban modeling requires extensive data processing with supporting technologies such as Geographic Information Systems (GIS) and relational database management systems (RDBMS) before the full value of the dataset can be exploited.

MassGIS, the State's Office of Geographic and Environmental Information, obtained access to the vehicle safety inspection records from the RMV for a “Climate Roadmap” project that details possible plans for significant reductions in GHG emissions for 2020-2050 in Massachusetts. MassGIS developed a statewide 250×250m grid cell layer to provide a standardized basis for spatially detailed analysis. That spatial unit of analysis is used in this study. A grid cell has an area of 15.4 acres, which is sufficiently small to capture spatial details and neighborhood effects. Compared with previous land use and transport studies for Metro Boston that are usually based on census geography, our study uses a much more fine-grained scale, especially for suburban areas where census spatial units are much larger than our grid cells. MassGIS geocoded each vehicle to an XY location approximating the owner's address, and tagged each VIN with the 250×250m grid cell ID containing that address using GIS tools[4]. MassGIS provided the VINs, XY locations, grid cell IDs and annual mileage estimates[5], to MIT for use in our research. Overall, 2.47 million private passenger vehicles are included in this dataset. Among them, 2.10 million vehicles (84.9%) have credible odometer readings. For the remaining 0.37 million vehicles, we know their place of garaging but do not have reliable odometer readings, either because the reported reading was determined to be in error (e.g., a lower odometer reading is recorded on a subsequent inspection date) or because two readings sufficiently far apart were not available, for example, for a brand new vehicle[6].

Population and household data from the 2000 Census block group data are allocated by MassGIS to those portions of Census block groups that are identified as residential by the State's 2000 land use dataset. Population and households are further assigned to grid cells based on the area and type of residential use in each grid cell. This is a significant refinement of block-group analysis of households and is repeatable for other states with good land use maps. In our empirical analysis, we compute VMT measures and built-environment characteristics at the grid cell level, and model their linkage using the demographic data as controls.

[4] People may change residential location between inspection periods. MassGIS addresses this problem as follows: (1) for VINs with same license plate number but different owner addresses in consecutive years, the pro-rated annual mileage is associated with the later address; (2) VINs with different plate numbers, suggesting a change of ownership during the observation period, are excluded from the study. While there could still be some spatial variation in assigning vehicles and vehicle mileage to grid cells, the impact of this mismatch on grid-cell-level aggregate VMT measures is likely to be limited.

[5] MassGIS compared the two recent vehicle inspection records for each private passenger vehicle, calculated the odometer reading difference, and pro-rated it based upon the time period between inspections so as to estimate annual mileage traveled.

[6] As a result, VMT measures of neighborhoods with a large number of new cars may be biased downwards. Biases may arise if vehicles without credible odometer readings are not distributed randomly in space and their mileages are significantly different from their neighbors.

3. METHODOLOGY

In this section, we present the methodology employed in this study, including model specification, variable generation and factor analysis.

3.1 Model Specifications

Equation 1 specifies the base model of the empirical analysis:

$$VMT_i = \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \tag{1}$$

where VMT_i is the zonal average VMT measures for grid cell i; BE_i is a vector of built-environment characteristics of grid cell i, and DEM_i is a vector of demographic characteristics of the block group that the centroid of grid cell i falls in[7].

To control for the interactions among the spatial units, we estimate both spatial-lag and spatial-error models (Anselin 2006). Spatial lag suggests a possible diffusion process - VMT of one zone is affected by the independent variables, for example, built-environment factors, of this zone as well as neighboring areas. With spatial lag in an OLS regression, the estimation result will be biased and inconsistent. Spatial error is indicative of omitted independent variables that are spatially correlated. With spatial error in an OLS regression, the estimation result will be inefficient. The spatial-lag model can be specified as:

$$VMT_i = \rho W_{VMT_i} + \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \tag{2}$$

where ρ is a spatial-lag correlation parameter, and ε is an Nx1 vector of i.i.d. standard normal errors. The spatial error model can be specified as:

$$\begin{aligned} VMT_i &= \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \\ \varepsilon_i &= \lambda W_{\varepsilon_i} + \mu_i \end{aligned} \tag{3}$$

[7] A list of BE and DEM variables/factors used in this study can be found in Table 1.

where λ is a spatial-error correlation parameter, and μ is an Nx1 vector of i.i.d. standard normal errors.

In Equations (2) and (3), W is the N×N matrix of spatial weights, which we developed assuming a constant spatial dependence among grid cells up to a maximum Euclidean distance of 750m.

3.2 VMT Variables

In this study, we explore the built-environment effects on three VMT measures: VMT per vehicle, VMT per household, and VMT per capita. We compute the VMT per vehicle for each grid cell based on vehicle-level annual mileage estimates from MassGIS. Some grid cells have very few vehicles. We apply the spatial interpolation function of GIS software to overcome issues related to sparse cells. For grid cells that have at least 12 vehicles with credible odometer readings (denoted as "good" cars), we assign the average annual mileage of all "good" cars to the grid cell. For other grid cells, we assign to them the inverse-distance-weighted average of 12 closest "good" annual mileages. We compute VMT per household (or per capita) for each grid cell by multiplying the estimated VMT per vehicle within the grid cell by the total number of vehicles within the grid cell and then divide by the number of households (or persons). These odometer-readings-based VMT estimates establish a baseline for tracking future changes in vehicle usage and associated energy consumptions and emissions for Metro Boston. Figure 2 plots the three VMT measures across grid cells in Metro Boston using quantile classification.

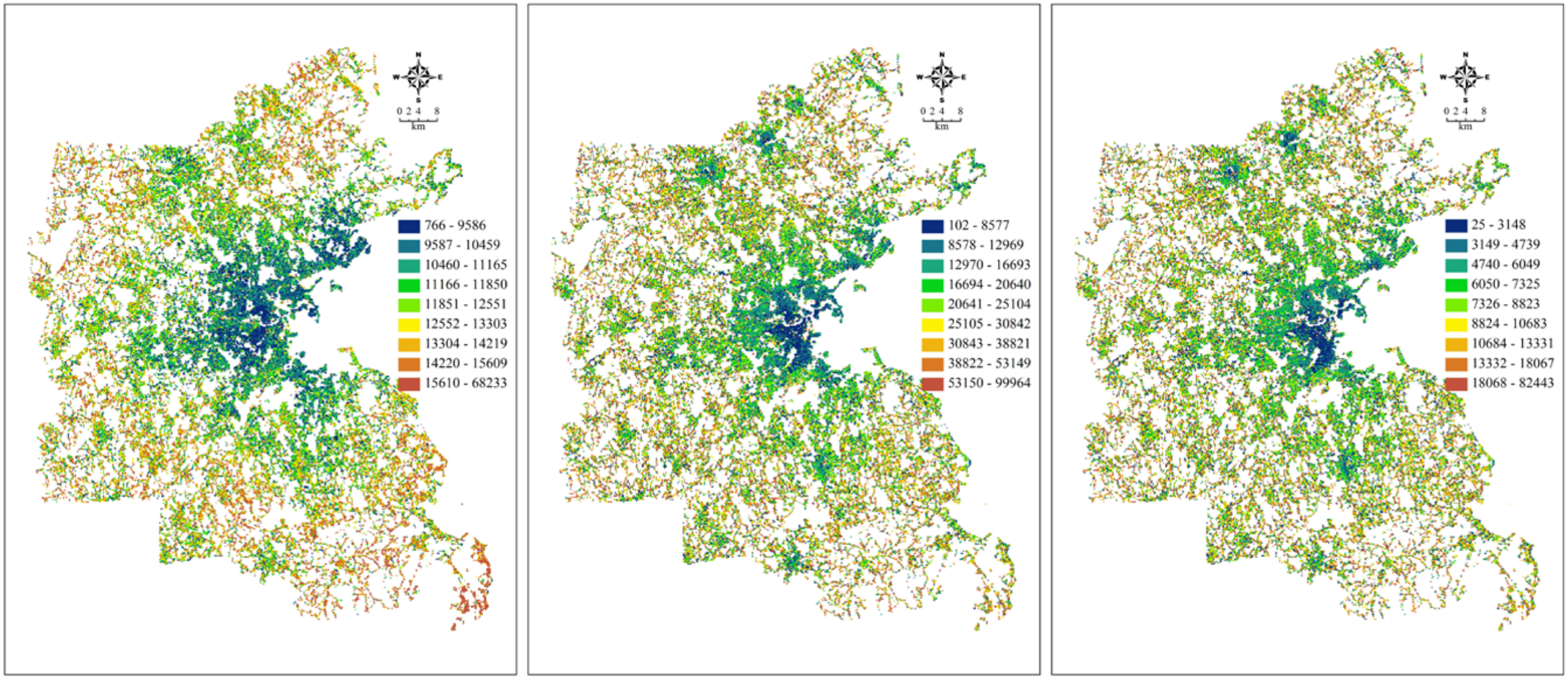
The overall spatial pattern is what researchers would expect: VMT is lower in grid cells near urban centers, but higher in suburban areas. It is also interesting to note: (a) large areas in the suburbs without vehicles or households; (b) significant variability within suburbs depending on whether the grid cell is near the town center; and (c) the difference in patterns between VMT per vehicle and VMT per household. The low VMT per vehicle in the inner city translates to even lower VMT per household due to the low car ownership level. Thus grid cells with low VMT per household are more concentrated in the urban core compared to grid cells with low VMT per vehicle.

Some measurement errors may arise in the variable generation from the georeferencing steps as well as from odometer recording errors. In this study, the locations of vehicles are street centerline locations estimated by MassGIS to be proximate to the home address using MassGIS address matching tools. Since street centerlines form the borders of block groups and house locations are offset 15m or more from street centerlines, any combination of measurement errors in basemap geometry or address geocoding could result in vehicles, households, or individuals being assigned to the wrong grid cell thereby resulting in a mismatch between households and vehicles. Such misassignments are likely to be random and not bias zonal VMT estimates for any particular grid cell. However, housing is less dense in the suburbs so suburban grid cells are more likely to be judged unreliable because they are estimated to have too few vehicles or households for the VMT per vehicle or VMT per household estimates to be reliable.

Furthermore, the allocation of households (or persons) to grid cells yields grid cells with only a sliver of residential area and, hence, only a fraction of one household (or person), which leads to superficially high VMT per household (or per capita) measures. The VMT per vehicle estimate does not depend on the estimated number of households (or persons) living in that grid cell. Hence it involves less measurement error than the VMT per household (or per capita) estimate.

Since we rely on odometer readings and vehicle counts from vehicle safety inspections to compute the VMT measures, our focus is limited to grid cells with at least one vehicle registered at the RMV (60,895 in total). To address the measurement error, we further eliminate grid cells estimated to have no households (or individuals) (2,870 grid cells) and grid cells with extreme VMT measures (annual zonal VMT per household less than 100 miles or greater than 100,000 miles). The final dataset for empirical analysis includes 53,188 grid cells that are estimated to have reliable estimates of VMT measures.

In Section 4, we fit regression models where the dependent variables are VMT per vehicle, VMT per household, and VMT per capita computed for the catchment area of each of the 53,188 grid cells. For these models, the catchment area of a grid cell is defined to be its nearest 9 grid cells. For example, the VMT per household is calculated as the total VMT divided by the total number of households in the catchment area of a grid cell. The catchment area can reduce the measurement errors due to the misassignment of vehicles, households, or individuals to grid cells, and errors related to the slivers of residential land by including the fractions of households/individuals in totals for their neighbors.



2(a) VMT per Vehicle   2(b) VMT per Household   2(c) VMT per Capita

Figure 2: VMT per Vehicle, VMT per Household and VMT per Capita across Grid Cells in Metro Boston

3.3 Built-Environment Variables

To better understand the relationship between the built environment and household vehicle usage, we need a better characterization of the built environment. In this study, we benefit from a set of built-environment datasets from MassGIS with exceptional spatial details, including the Dun and Bradstreet business location database and MassGIS records of institutional location, land use, road networks, and grid cell level counts of population and households. We compute 27 variables at the 250m×250m grid cell level along multiple dimensions to characterize the built environment, as shown in Table 1.

The eight variables of distances to major non-work destinations, including shopping malls, grocery stores, schools, hardware stores, restaurants, churches, dentists, and gyms, are from MassGIS. MassGIS computed distances to a variety of non-work destinations at the grid cell level using GIS tools. We select the eight most frequent destination types from the 29 types in the 2001 National Household Transportation Survey. For built-environment variables that rely on a definition of neighborhood, such as density and land-use mix, we define a catchment area (neighborhood) for each grid cell as the 3x3 nearest grid cells, compute the variable of interest for the catchment area, and assign the value to the grid cell in the middle. The catchment area has a size that is close to the transportation impact area, which is conventionally defined as a circle with a 1⁄4-mi radius, a figure that has been backed by behavioral and empirical research. The adoption of the grid cell and catchment area approach could mitigate the potential biases and inconsistency caused by the MAUP as discussed in Diao and Ferreira (2010).

Many built-environment variables tend to be closely correlated. For example, relatively dense neighborhoods tend to have a greater variety of land uses, smaller blocks, and so on. A regression model with highly-correlated variables is likely to result in numerous insignificant or incorrectly-signed coefficients. To deal with the multicollinearity, we perform a principle component analysis with Varimax rotation to reduce the 27 built-environment variables to a small set of factors. The top 5 factors with initial eigenvalues greater than 1 explain 69.8% of variance in original variables. In other words, there is only a 30% loss in information incurred by the 82% reduction in the number of built-environment variables from 27 to 5. The 5 factors are included in regression models.

Factor 1 has high loadings on variables measuring distance to non-work destinations and land-use mix, and therefore describes primarily “distance to non-work destinations”. Grid cells with higher scores in factor 1 tend to have longer distance to non-work destinations, and thus are hypothesized to have higher VMT (others factors held constant). Factor 2 places the highest weights on street network layout and population density. We label it as “connectivity”[8]. Good connectivity can improve the connection of people and places and shorten local trips, thereby reducing vehicle usage. Factor 3 describes the difficulty of accessing transit systems and jobs, with positively high loadings on distance to transit variables and negatively high loading on job accessibility. We expect factor 3 to be positively associated with VMT. Factor 4 represents the degree of auto dominance, that is, the extent to which automobile movement is facilitated in the locality, for example, by having wider roads, higher speed limits, etc. It could decrease travel costs of the auto mode, thus increasing vehicle usage. The fifth factor “walkability” describes the pedestrian environment, which can reduce the travel costs of walking, thus decreasing VMT. Compared with grid cells in the suburbs, grid cells in urban centers have better accessibility to non-work destinations, jobs, and transit systems, better connectivity, and better pedestrian

[8] In built environment research, good connectivity often means a street network provides multiple routes and connections between origin and destinations. In this study, our factor analysis collapsed population density and multiple street network layout indicators into one factor. Since most of the variables with high factor loadings in this factor describe street network layout, we label it “connectivity ”.

environment as expected. Grid cells with higher scores in the “auto dominance” factor tend to be located along major transportation corridors.

3.4 Demographic Variables

In this study, we select 12 demographic variables at the block group level to control for the zonal difference of population as shown in Table 1. Ideally, we should compute demographic variables at the grid cell level, but because of data limitations, we assign each grid cell the value of the block group that contains its centroid. Since block groups can be much larger than grid cells outside the inner cities, some spatial autocorrelation is built in.

Similarly, we also apply a principle component analysis with Varimax rotation to the 12 demographic variables and extract from them 3 demographic factors: wealth, children, and working status. Factor 1 can be seen as an indicator of wealth. Block groups with higher values in Factor 2 tend to have more children and bigger household size. Factor 3 is related to residents’ working status. The three factors explain 71.6% of the variance in the original variables.

Factor loadings for built-environment and demographic factors are presented in Table 1. Table 2 shows the descriptive statistics of variables in the regression models.

Table 1: Factor Loadings on Factors

	BE Variables	BE factor 1 Distance to non-work destinations	BE factor 2 Connectivity	BE factor 3 Inaccessibility to transit and jobs	BE factor 4 Auto dominance	Be factor 5 Walkability
1	Distance to restaurant	0.784				
2	Distance to mall	0.764				
3	Distance to hardware store	0.746				
4	Distance to grocery	0.733				
5	Distance to dentist	0.688		0.398		
6	Distance to gym	0.676				
7	Distance to church	0.674				
8	Distance to school	0.645				
9	Land-use mix	-0.480				
10	Density of 4-way intersections		0.872			
11	Intersection density		0.849			
12	Density of 3-way intersections		0.809			
13	Population density		0.785			
14	Road density	-0.353	0.765			
15	Percent of 4-way intersections		0.609			
16	Distance to bus stop			0.833		
17	Distance to commuter rail station			0.810		
18	Distance to subway station			0.801		
19	Distance to MBTA parking lot			0.775		
20	Job accessibility		0.486	-0.636		
21	Percent of roads with access control				0.910	
22	Average road width				0.875	
23	Percent of roads with 30+ speed limit				0.856	
24	Distance to highway exit				-0.362	
25	Percent of roads with sidewalks					0.910
26	Percent of roads with curbs					0.908
27	Average sidewalk width		0.583			0.602

	DEM Variables	DEM factor1 Wealth	DEM factor 2 Children	DEM factor 3 Working Status
1	Percent of population below poverty level	-0.863		
2	Percent of owner-occupied housing units	0.818	0.386	
3	Percent of population with at least 13 years of schooling	0.817		
4	Median household income	0.812		
5	Percent of population that is white	0.796		
6	Per capita income	0.707		
7	Unemployment rate	-0.613		
8	Percent of households with less than 3 members		-0.909	
9	Percent of population that are enrolled in elementary high school		0.869	
10	Percent of population under 5		0.728	
11	Percent of population 65 years old and over			-0.856
12	Percent of population 16+ years in labor force	0.427		0.793

* We suppress factor loadings with absolute value less than 0.35 for interpretation convenience.

Table 2: Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
VMT per vehicle	53,188	12,547.1	1,888.9	1,880.7	42,778.0
VMT per household	53,188	25,641.4	11,800.3	543.2	99,715.8
VMT per capita	53,188	8,903.0	3,764.0	84.2	54,154.5
BE factor. 1: distance to non-work destinations	53,188	-0.239	0.864	-2.590	3.980
BE factor 2: connectivity	53,188	0.422	1.167	-1.640	11.120
BE factor 3: inaccessibility to transit and jobs	53,188	-0.104	0.970	-2.270	4.580
BE factor 4: auto dominance	53,188	-0.078	0.611	-1.200	6.400
BE factor 5: walkability	53,188	0.078	0.916	-2.660	4.000
DEM factor 1: wealth	53,188	0.565	0.650	-4.150	2.580
DEM factor 2: children	53,188	0.412	0.760	-3.320	3.790
DEM factor 3: working status	53,188	0.098	0.858	-6.920	4.100

4. EMPIRICAL ANALYSIS

In this section, we present the modeling results of the empirical analysis. Depending upon the selection among 3 dependent variables and 3 model specifications, we estimate the following 9 models:

(1) OLS model for VMT per vehicle;

(2) OLS model for VMT per household;

(3) OLS model for VMT per capita;

(4) Spatial-lag model for VMT per vehicle;

(5) Spatial-lag model for VMT per household;

(6) Spatial-lag model for VMT per capita;

(7) Spatial-error model for VMT per vehicle;

(8) Spatial-error model for VMT per household; and

(9) Spatial-error model for VMT per capita.

We estimate the spatial-lag and spatial-error models with GeoDa 0.9.5 software. Table 3 summarizes statistics for the regression models. Residual tests indicate that the error terms of the OLS models exhibit significant spatial autocorrelation. Moreover, both the simple Lagrange multiplier tests for omitted spatially-lagged dependent variables (LM-lag) and error dependence (LM-error) are statistically significant, indicating the possible existence of both spatial-lag and spatial-error types of spatial autocorrelation.

Robust Lagrange multiplier tests for spatial-lag and spatial-error specifications are employed to identify the major source of spatial autocorrelation. Both the test for error dependence in the possible presence of a missing lagged dependent variable (robust LM-error), and the test for a missing lagged dependent variables in the possible presence of spatially-correlated error term (robust LM-lag) are statistically significant. However, the robust LM-error test rejects the null hypothesis at a higher level of significance, favoring the spatial-error model. The R-squared and log-likelihood statistics also support this conclusion, indicating that the spatial-error model generally has a better fit to the data than the corresponding spatial-lag model and OLS model. These results suggest that some spatially correlated variables, other than the built-environment and demographic factors captured in this study are the major source of spatial autocorrelation. Table 4 presents the estimation results of the three models using the spatial-error specification.

The Breusch-Pagan tests confirm the existence of heteroskedasticity before and after the spatial autocorrelation are controlled for, with similar values (both significant at the 0.01 level). This is no surprise since grid cells in the suburbs are less dense and have systematically higher VMT per vehicle (and per household or per capita). Allowing the error terms to be spatially correlated is one way of accounting for the spatial effect and the spatial-error model did improve the model fit, but it didn't make the spatial effects go away.

As shown in Table 4, most coefficients for demographic factors are statistically significant. One interesting finding is that higher "wealthy" levels are associated with **lower** VMT ***per vehicle***, but **higher** VMT ***per household*** and VMT ***per capita***, which shows that households in wealthier neighborhoods tend to own more cars but drive each car a little less compared to households in other neighborhoods (after controlling for other factors). The number of children in the household is positively associated with VMT per household, presumably because of child-related non-work trips. But the effects of children on VMT per vehicle and VMT per capita are insignificant. One possible explanation is that households tend to buy more vehicles as household size grows, but the usage of each vehicle does not change significantly. Factor 3 can be seen as a proxy for the percentage of population that is working. This factor is positively associated with all three VMT variables, presumably due to the commuting trips.

Table 3: Estimation Summary

	VMT per Vehicle			VMT per Household			VMT per Capita		
	OLS	Spatial Lag	Spatial Error	OLS	Spatial Lag	Spatial Error	OLS	Spatial Lag	Spatial Error
Observations	53,188	53,188	53,188	53,188	53,188	53,188	53,188	53,188	53,188
R-squared	0.456	0.718	0.736	0.344	0.550	0.552	0.261	0.487	0.492
Log Likelihood	-460,512	-445,191	-443,975	-562,936	-554,633	-554,704	-505,331	-497,344	-497,252
Test	Statistic	p-value		Statistic	p-value		Statistic	p-value	
LM--Lag	69,021.6	0.00		35,686.8	0.00		34,735.9	0.00	
LM--Error	86,887.6	0.00		36,260.5	0.00		35,656.0	0.00	
Robust LM--Lag	684.6	0.00		825.2	0.00		328.3	0.00	
Robust LM--Error	18,550.6	0.00		1,398.9	0.00		1,248.3	0.00	

Table 4: Estimation Results of Spatial-Error Models

	VMT per Vehicle			VMT per Household			VMT per Capita		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
Built-Environment Factors									
Distance to non-work destinations	381.4	16.30	**	2,581.9	17.63	**	461.3	9.42	**
Connectivity	-248.1	-20.73	**	-1,826.0	-23.37	**	-454.3	-17.31	**
Inaccessibility to transit & jobs	1,033.9	32.56	**	4,813.0	29.22	**	1,638.2	30.04	**
Auto dominance	42.2	3.58	**	676.1	7.36	**	293.1	9.43	**
Walkability	-9.2	-0.89		-1,165.1	-15.38	**	-446.2	-17.44	**
Demographic Factors									
Wealth	-50.4	-3.06	**	803.2	6.32	**	316.2	7.35	**
Children	0.2	0.02		632.8	7.05	**	-26.3	-0.86	
Working status	29.5	3.50	**	155.0	2.32	*	59.5	2.62	**
Lambda	0.89	316.31	**	0.80	197.80	**	0.80	192.37	**
Constant	12,880.3	329.93	**	27,505.1	135.58	**	9,434.0	140.08	**

* and ** denote coefficient significant at the 0.05 and 0.01 level respectively.

After controlling for demographic factors, we find that built-environment factors are indeed important predicators of vehicle usage at the grid cell level, with smart-growth-type neighborhoods associated with less vehicle usage than sprawl-type neighborhoods. The coefficients for the "distance to non-work destination" factor in the three models are positive and significant at the 0.01 level, suggesting that the spatial distribution of non-work activities is significantly associated with vehicle usage. As the distance to non-work destinations increases, VMT per vehicle, VMT per household, and VMT per capita all increase. The negative sign of the "connectivity" factor in all three models suggests that connectivity – an indicator of high-density, grid-type neighborhoods – is associated with reduced vehicle usage. Accessibility to the transit system and job centers plays a critical role in explaining the intra-urban variation of VMT, as reflected in the positive and significant coefficients of the "inaccessibility to transit and jobs" factor in all three models. The coefficients of the "auto dominance" factor are positive and significant in the three models. This suggests that auto-friendly environments are associated with higher VMT. As revealed by the estimated coefficients of the "walkability" factor, a good pedestrian environment is associated with lower VMT per household and VMT per capita, while its effect on VMT per vehicle is insignificant. Hence, the "walkability" factor tends to influence VMT by reducing the number of vehicles purchased.

By comparing the coefficients of the demographic and built-environment factors, we find that built-environment factors have a higher predictive power[9] on VMT than demographic factors. Table 5 and Figure 3 present the change in annual VMT per vehicle, per household, and per capita due to one standard deviation increase in individual factors. As shown in Figure 3, accessibility to work and non-work destinations, connectivity, and transit accessibility generally

---

[9] Since the independent variables are already normalized factor scores, the estimated coefficients are standardized and can be directly compared.

make much higher contributions to the models than other factors[10]. The contributions are relatively large for the VMT per household measure, where the average VMT per household at grid cell level for the study area is about 25,641 miles.

Table 5: Change in VMT Measures Due to One Standard Deviation Increase in Factors

	VMT per Vehicle	VMT per Household	VMT per Capita
Built Environment Factors			
Distance to non-work destinations	329.7	2,231.7	398.7
Connectivity	-289.4	-2,130.1	-530.0
Inaccessibility to transit and jobs	1,002.6	4,667.5	1,588.7
Auto dominance	25.8	413.0	179.1
Walkability	-8.4	-1,067.6	-408.8
Demographic Factors			
Wealth	-32.8	522.1	205.6
Children	0.1	481.0	-20.0
Working Status	25.3	133.1	51.0

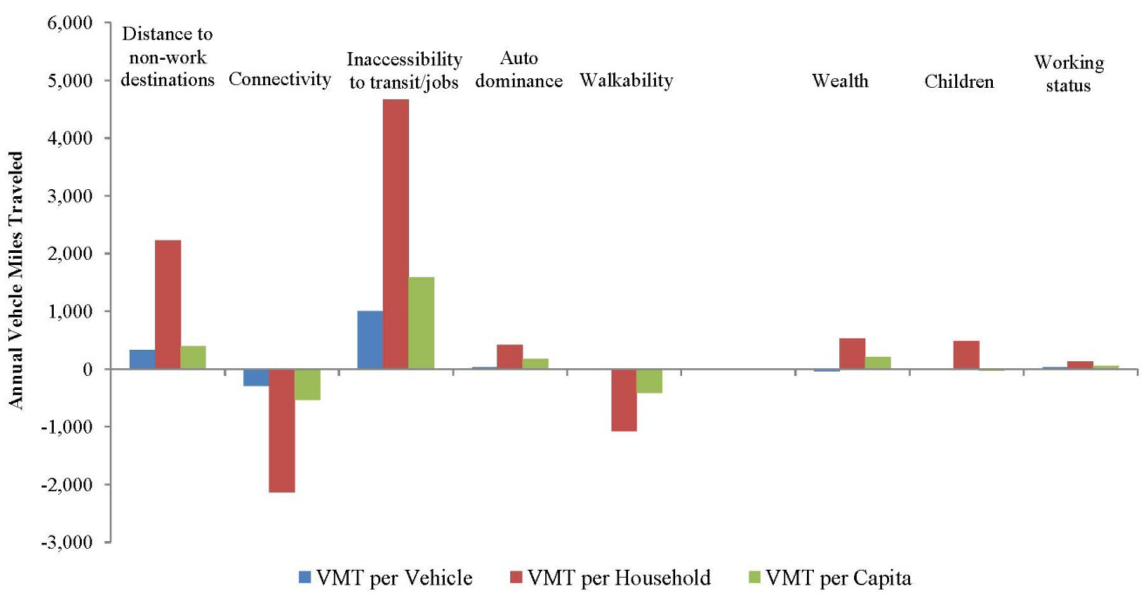


Figure 3: Contributions of Factors to the Model

To understand how the linkage between household vehicle usage and residential locations varies across different types of communities, we calibrate stratified spatial-error models for grid

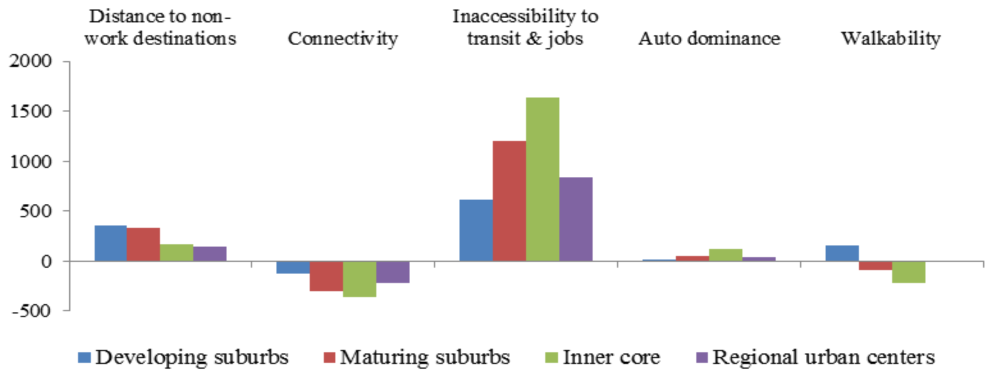
[10] One exception is that the distance to non-work destination factors has a slightly lower effect than the walkability factor in the VMT per capita model.

cells in four types of communities: developing suburbs, maturing suburbs, inner core, and regional urban centers. The community type is based on the classification of the 164 municipalities in the region by the regional planning agency, Metropolitan Area Planning Council (See Figure 1). The developing suburbs have above average growth rate and plenty of vacant land available for development. Some have strong town centers and moderate density neighborhoods, others are more rural. The maturing suburbs are mostly moderate density neighborhoods with average growth rates and a dwindling supply of unprotected developable land. The inner core is characterized by high-density neighborhoods, multifamily housing, large immigrant population, and below average growth rate. The regional urban centers are urban neighborhoods with large immigrant communities and below average growth rates. Some still have large amounts of developable land[11].

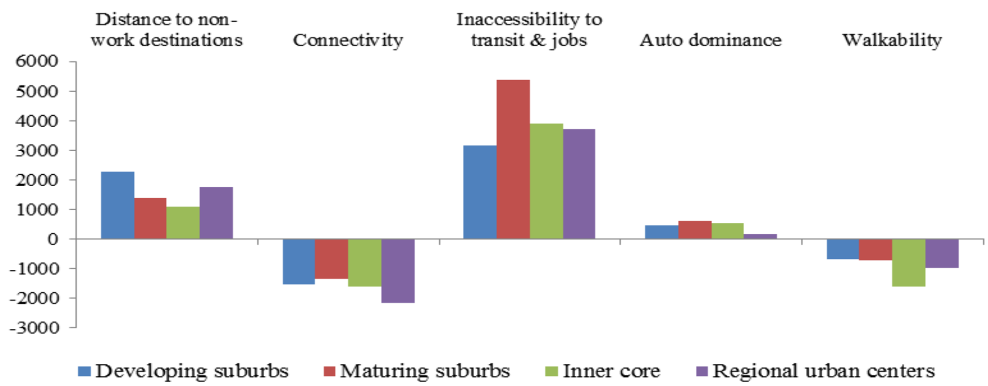
Based on the estimation results, the annual VMT increases corresponding to one standard deviation increases of built-environment factors are plotted by community types in Figure 4. The results are generally consistent with the pooled model, with similar estimated coefficients for the built-environment and demographic factors, but also display some notable differences across community types. For example, some built-environment characteristics, such as distance to non-work destinations and connectivity play a less important role in explaining the variations in vehicle usage in the inner core compared to in other types of communities, while walkability is relatively more important for the inner core. Since inner-city locations are close to many non-work destinations, the reduced coefficient is no surprise. The walkability results are less obvious. The sign is always negative and significant for VMT per household and per capita. However the per vehicle effects are mixed and significantly positive in the developing suburbs but

[11] Source: Metropolitan Area Planning Council

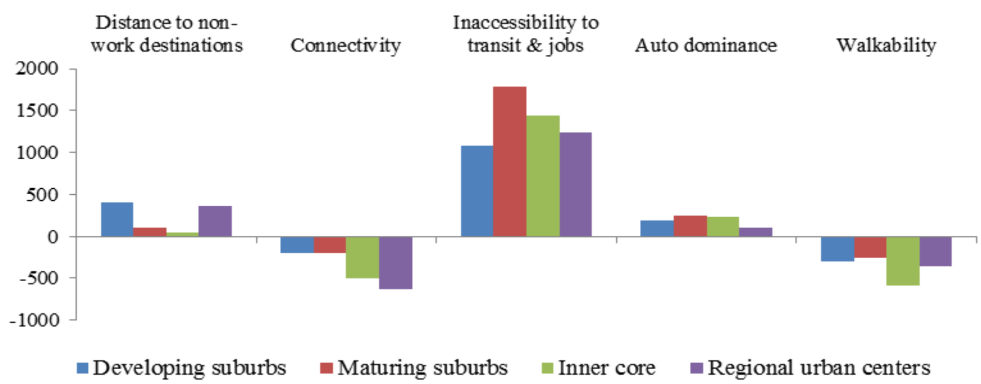
significantly negative in the inner city. One possible explanation is that walkability does reduce VMT per household. In the inner city, the reduction comes from driving a household's one car less often, but in the suburb it comes from doing without a second (or third) car. Those few parts of the developing suburbs that are walkable may be where households own fewer cars, but drive the one car they have a little more.



4(a) VMT per vehicle



4(b) VMT per household



4(c) VMT per Capita

Figure 4: Annual VMT increase due to One Standard Deviation Increase of BE factors by Community Types

5. CONCLUSIONS

In this study, we examine the relationship between household vehicle usage and their residential locations in the Boston Metropolitan Area with VMT measures derived from a georeferenced administrative dataset - the odometer readings from annual safety inspection for all private passenger vehicles registered in Metro Boston. These VMT measures are compared with spatially-detailed built-environment measures and demographic measures developed from census block group data allocated to residential grid cells. Spatial regression models are calibrated to control for the potential spatial autocorrelation.

Our models perform well in explaining the intra-urban variations of VMT within Metro Boston at the smoothed 250×250m grid cells level. The regression results reveal that built-environment factors, particularly accessibility to work and non-work destinations, connectivity and transit accessibility, have stronger association with VMT than demographic factors. In part, this may be due to measurement issues since the demographic variables are computed at the more aggregated block group level and their effects could be underestimated. However, in most studies using travel survey data, the bias is in the other direction – the individual demographic characteristics are known, but the built-environment factors come from data aggregated at census tract or block group scale. Many of these studies find small built-environment effect (e.g., Brownstone and Golob 2009; Bento et al. 2005). But their built-environment effect is likely to be biased downwards (relative to the demographic effect) since they mix aggregate built-environment measures with individual household characteristics.

The results also highlight important differences between VMT per vehicle, per person, and per household. For example, Figure 2 visualizes some of the spatial differences and Table 4 shows that the demographic factor interpreted as ‘wealth’ was negatively correlated with VMT per vehicle but had a positive (and much stronger) correlation with VMT per person and per household. Areas with higher ‘wealth’ tend to have considerably more cars per household. Also, all the built-environment effects were disproportionally larger for the per household model indicating that the combination of vehicle ownership and usage enhances the built-environment effects at the household level.

As land use and transportation modeling shifts from trip-based mesoscopic models to activity-based microsimulations, it will be increasingly important to address built-environment effects in detail. This study contributes to the literature by demonstrating fine-grained spatial variation in the VMT correlations, and by obtaining effects that are considerably larger than in studies where the built-environment measures are more aggregated.

Finding a strong association between the built environment and travel patterns is not the same as showing that a change in the built environment will lead to a change in travel behavior (Handy 1996). Self-selection bias could explain some of the observed correlation, and one would expect a mix of selection bias and geographic effects. The strength and consistency of the findings provide some support for those smart-growth policies that advocate increasing accessibility to destinations, creating traditional-type high-density, mixed-use neighborhoods, and improving transit accessibility. At a minimum, our results suggest that built-environment characteristics matter at a fine-grained level of detail and may be relatively more important than many previous studies suggest. The VMT effects in Figure 3 associate sizeable VMT differences (1-5k increases in annual mileage per-household) with a one-standard deviation change in built-environment factors. This result suggests that regulating urban form could be a meaningful way strategy for reducing transport GHG reduction[12]. It is true that a particular neighborhood cannot easily be converted to be half-a-standard deviation higher or lower on its built-environment factor. While it is hard to change the built environment of any given neighborhood, that also means that any built-environment effects are going to persist for a long time and tend to lock in the degree of car usage that is necessarily associated with a type of neighborhood.

The built-environment effects are reasonably consistent in sign and magnitude across community type, but they do vary somewhat, suggesting that different types of neighborhoods may demand different smart-growth policies to reduce vehicle usage effectively. Furthermore, the study can be easily replicated to conduct time series analyses and to build tractable indicators for monitoring the performance of metro areas in reducing transport GHG emissions and tracking longitudinal changes in land use-transportation interconnections.

This study also has implications for urban modeling by revealing the opportunities brought about by georeferenced administrative data. These datasets are collected regularly by various agencies for management purposes. Using vehicle safety inspection records as an example, this study demonstrates that georeferenced administrative data can be quite useful, but transforming the raw data into useful formats requires intensive data processing with support of technologies such as GIS and DBMS tools. Most states already have GIS offices that could routinely build such derived data layers. In the future, as parcel level data and address matching tools keep improving, better vehicle and household counts will become available thereby reducing the measurement errors associated with the vehicle safety inspection data. Meanwhile, researchers have been experimenting with other types of administrative datasets in urban modeling, such as mobile phone traces (Calabrese et al. 2013), transit card transactions (Wilson

[12] We used this example of effects instead of reporting elasticities since the elasticity of annual mileage with respect to changes in built environment is not constant.

et al. 2009), and GPS traces (Grengs et al. 2008). Building indicators and calibrating urban models using such administrative data can complement the high expense of traditional surveys and enable improved monitoring and modeling of metropolitan areas at a spatially-detailed scale.

This study can be extended along multiple directions, for example:

(1) parsing VINs to obtain EPA fuel consumption estimates so that the built-environment effects can be combined with vehicle ownership patterns to estimate energy consumption patterns and GHG emissions.

(2) employing structural equation models to investigate the causal relationships among key variables, such as the built environment, automobile ownership, and travel behavior;

(3) extending the analysis to other metropolitan areas, and

(4) improving the demographic indicators by using parcel and housing unit level indicators derived from assessing, voting, and other local government datasets.

ACKNOWLEDGEMENT

We acknowledge the foresight and geoprocessing skill of MassGIS in developing some of the unique datasets used in this research. We also acknowledge the helpful comments and suggestions of the referees and the partial support of the US Department of Transportation Region One University Transportation Center through grant MITR21-4, and the Singapore National Research Foundation through the "Future Urban Mobility" program of the Singapore-MIT Alliance for Research and Technology.

REFERENCE

1. Anselin, L., 2006. Spatial Econometrics. In T.C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Basingstoke, Palgrave Macmillan.
2. Bento, A., Cropper, M., Mobarak, A.M., Vinha, K., 2005. The effect of urban spatial structure on travel demand in the United States. *Review of Economics and Statistics* 87(3): 466-478.
3. Boarnet, M., Crane, R., 2001.The influence of land use on travel behavior: Specification and estimation strategies. *Transportation Research Part A* 35: 823-845.
4. Brownstone, D., 2008. Key relationships between the built environment and VMT. Unpublished Paper prepared for the Transportation Research Board and the Division on Engineering and Physical Sciences.
5. Brownstone, D., Golob T.F., 2009. The Impact of Residential Density on Vehicle Usage and Energy Consumption. *Journal of Urban Economics* 65: 91-98.
6. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti. C., 2013. Understanding individual mobility patterns with urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies* 26: 301-313.
7. Diao, M., Ferreira, J., 2010. Residential property values and the built environment: Empirical study in the Boston, Massachusetts, Metropolitan Area. *Transportation Research Record: Journal of the Transportation Research Board* 2174: 138-147.
8. Ewing, R., Cervero, R., 2001. Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board* 1780: 87-113.
9. Ferreira, J., Diao, M., Xu, J., 2013. Estimating the vehicle-miles-traveled implications of alternative metropolitan growth scenarios – A Boston example, *Transaction in GIS* 17 (5): 645-660.
10. Grengs, J., Wang, X., Kostyniuk, L., 2008. Using GPS data to understand driving behavior. *Journal of Urban Technology* 15(2): 33-53.
11. Handy, S., 1996. Methodologies for exploring the link between urban form and travel behavior. *Transportation Research Part D* 1(2): 151-165.
12. Handy, S., Cao, X., Mokhtarian, P., 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D* 10: 427-444.
13. Holtzclaw, J., 1994, Using residential patterns and transit to decrease auto dependence and costs. Natural Resources Defense Council, San Francisco, and California Home Energy Efficiency Rating Systems, Costa Mesa, California.
14. Holtzclaw, J., Clear R, Dittmar H., Goldstein D., Hass P., 2002, Location efficiency: Neighborhood and socioeconomic characteristics determine auto ownership and use – Studies in Chicago, Los Angeles and San Francisco. *Transportation Planning and Technology* 25: 1–27.
15. International Energy Agency (IEA), 2004. *The IEA/SMP Transport Spreadsheet Model,* developed for the World Business Council for Sustainable Development Sustainable Mobility Project.
16. Lindsey, M., Schofer, J., Durango-Cohen, P., Gray, K., 2011. The effect of residential location on vehicle miles of travel, energy consumption and greenhouse gas emissions: Chicago case study. *Transportation Research Part D* (16): 1–9.
17. Miller, E.J., Ibrahim, A., 1998. Urban form and vehicular travel: Some empirical findings. *Transportation Research Record: Journal of the Transportation Research Board* 1617: 18-27.
18. Price, L., de la Rue du Can, S., Sinton, J., Worrell, E., Nan, Z., Sathaye, J., Levine, M., 2006. *Sectoral Trends in Global Energy Use and Greenhouse Gas Emissions LBNL-56144*. Ernest Orlando Berkeley National Laboratory, Environmental Energy Technologies Division, Berkeley, CA, July 2006.

19. Schipper, M., Moorhead, V., 2000. Odometer versus self-reported estimates of vehicle miles traveled. http://www.eia.doe.gov/emeu/consumptionbriefs/transportation/vmt/vmt.html, Release date: Aug. 3, 2000.
20. Wilson, N., Zhao, J., Rahbee, A., 2009. The potential impact of automated data collection systems on urban public transport planning. In N. Wilson and A. Nuzzolo (Eds.) *Schedule-Based Modeling of Transportation Networks: Theory and Applications*. Springer.
21. Yang, J., 2008. Policy implications of excess commuting: examining the impacts of changes in US metropolitan spatial structure. *Urban Studies* 45(2): 391–405.
22. Zhang, M., 2004. The role of land use in travel mode choice: Evidence from Boston and Hong Kong. *Journey of the American Planning Association* 70(3): 344-360.